mini Manuel de

Génétique

2e édition



Jean-Michel Petit
Sébastien Arico
Raymond Julien









DUNOD

www.sba-medecine.com





de Génétique

Cours + QCM/QROC

Jean-Michel Petit
Maître de conférences à l'université de Limoges

Sébastien Arico
Directeur R&D "Ingenomix" Limoges

Raymond Julien
Professeur émérite à l'université de Limoges

2e édition



DUNOD

Les illustrations de cet ouvrage ont été réalisées par Sébastien ARICO.

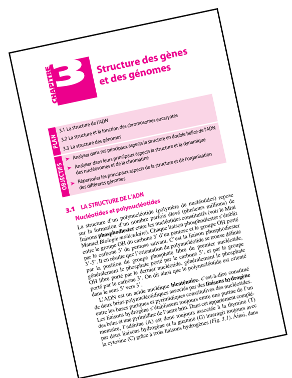
Directeur d'ouvrage
Raymond JULIEN

ISBN 978-2-10-055930-5



Comment utiliser le Mini-Manuel

La page d'entrée de chapitre



CHAPITRE 3

Structure des gènes et des génomes

PLAN

3.1 La structure de l'ADN
3.2 La structure et la fonction des chromosomes eucaryotes
3.3 La structure des génomes

OBJECTIFS

- Analyser dans ses principaux aspects la structure en double hélice de l'ADN
- Analyser dans leurs principaux aspects la structure et la dynamique des nucléosomes et de la chromatine
- Répertorier les principaux aspects de la structure et de l'organisation des différents génomes

3.1 LA STRUCTURE DE L'ADN

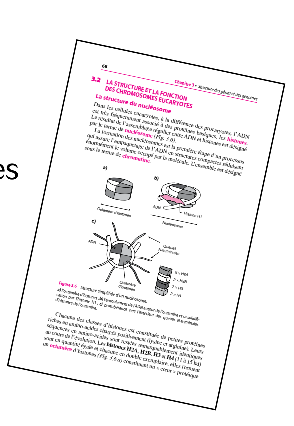
Nucléotides et polynucléotides

La structure d'un polynucléotide (polymère de nucléotides) repose sur la formation d'un nombre parfois élevé (plusieurs millions) de liaisons **phosphodiester** entre les nucléotides constitutifs (voir le Mini Manuel *Biologie moléculaire*). Chaque liaison phosphodiester s'établit entre le groupe OH du carbone 3' d'un pentose et le groupe OH porté par le carbone 5' du pentose suivant. C'est la liaison phosphodiester 3'-5'. Il en résulte que l'orientation du polynucléotide se trouve définie par la position du groupe phosphate libre du premier nucléotide, généralement le phosphate porté par le carbone 5', et par le groupe OH libre porté par le dernier nucléotide, généralement le phosphate porté par le carbone 3'. On dit ainsi que le polynucléotide est orienté dans le sens 5' vers 3'.

L'ADN est un acide nucléique **bicaténaire**, c'est-à-dire constitué de deux brins polynucléotidiques associés par des **liaisons hydrogène** entre les bases puriques et pyrimidiques constitutives des nucléotides. Les liaisons hydrogène s'établissent toujours entre une purine de l'un des brins et une pyrimidine de l'autre brin. Dans cet appariement complémentaire, l'adénine (A) est donc toujours associée à la thymine (T) par deux liaisons hydrogène et la guanine (G) interagit toujours avec la cytosine (C) grâce à trois liaisons hydrogènes (*Fig. 3.1*). Ainsi, dans

Elle donne le plan du cours, ainsi qu'un rappel des objectifs pédagogiques du chapitre

Le cours



68

Chapitre 3 • Structure des gènes et des génomes

3.2 LA STRUCTURE ET LA FONCTION DES CHROMOSOMES EUCARYOTES

La structure du nucléosome

Dans les cellules eucaryotes, à la différence des procaryotes, l'ADN est très fréquemment associé à des protéines basiques, les **histones**. Le résultat de l'assemblage régulier entre ADN et histones est désigné par le terme de **nucléosome** *(Fig. 3.6)*.

La formation des nucléosomes est la première étape d'un processus qui assure l'empaquetage de l'ADN en structures compactes réduisant énormément le volume occupé par la molécule. L'ensemble est désigné sous le terme de **chromatine**.

Figure 3.6 Structure simplifiée d'un nucléosome.
a) l'octamère d'histones ; b) l'enroulement de l'ADN autour de l'octamère et sa solidification par l'histone H1 ; c) protubérance vers l'extérieur des queues N-terminales d'histones de l'octamère.

Chacune des classes d'histones est constituée de petites protéines riches en amino-acides chargés positivement (lysine et arginine). Leurs séquences en amino-acides sont restées remarquablement identiques au cours de l'évolution. Les **histones H2A**, **H2B**, **H3** et **H4** (11 à 15 kd) sont en quantité égale et chacune en double exemplaire, elles forment un **octamère** d'histones *(Fig. 3.6 a)* constituant un « cœur » protéique

Le cours, concis et structuré, expose les notions importantes du programme

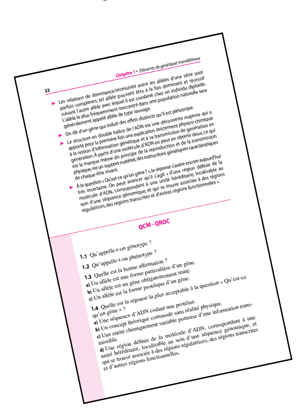
Les rubriques

 Une erreur à éviter

 Un peu de méthode

 Un exemple pour comprendre

 Les points clés à retenir



22

Chapitre 1 • Éléments de génétique mendélienne

- Les relations de dominance/récessivité entre les allèles d'une série sont parfois complexes, tel allèle pouvant être à la fois dominant et récessif suivant l'autre allèle avec lequel il est combiné chez un individu diploïde. L'allèle le plus fréquemment rencontré dans une population naturelle sera généralement appelé allèle de type sauvage.
- On dit d'un gène qui induit des effets distincts qu'il est pléiotrope.
- La structure en double hélice de l'ADN est une découverte majeure qui a apporté pour la première fois une explication strictement physico-chimique à la notion d'information génétique et à sa transmission de génération en génération. À partir d'une molécule d'ADN on peut en obtenir deux, ce qui est la marque même du principe de la reproduction et de la transmission physique, *via* un support matériel, des instructions génétiques caractéristiques de chaque être vivant.
- À la question « Qu'est-ce qu'un gène ? », la réponse s'avère encore aujourd'hui très incertaine. On peut avancer qu'il s'agit « d'une région définie de la molécule d'ADN, correspondant à une unité héréditaire, localisable au sein d'une séquence génomique, et qui se trouve associée à des régions régulatrices, des régions transcrites et d'autres régions fonctionnelles ».

QCM - QROC

1.1 Qu'appelle-t-on génotype ?

1.2 Qu'appelle-t-on phénotype ?

1.3 Quelle est la bonne affirmation ?
a) Un allèle est une forme particulière d'un gène.
b) Un allèle est un gène obligatoirement muté.
c) Un allèle est la forme protéique d'un gène.

1.4 Quelle est la réponse la plus acceptable à la question « Qu'est-ce qu'un gène » ?
a) Une séquence d'ADN codant une protéine.
b) Un concept théorique commode sans réalité physique.
c) Une entité chimiquement variable porteuse d'une information transmissible.
d) Une région définie de la molécule d'ADN, correspondant à une unité héréditaire, localisable au sein d'une séquence génomique, et qui se trouve associée à des régions régulatrices, des régions transcrites et d'autres régions fonctionnelles.

Les exercices, QCM ou QROC

Ils sont proposés en fin de chapitre, avec leur solutions, pour se tester tout au long de l'année.


sba-medecine
La communauté médicale des étudiants en médecine

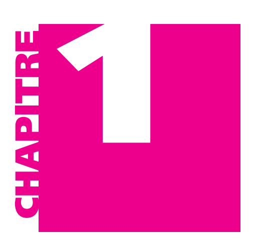


Table des matières


sba-medecine
La communauté médicale des étudiants en médecine


sba-medecine
La communauté médicale des étudiants en médecine





Éléments de génétique mendélienne

PLAN

OBJECTIFS

- Réviser les éléments de base de l'analyse mendélienne
- Approfondir la notion de caractère héréditaire
- Examiner la relation entre génotype et phénotype
- Réfléchir à la notion de gène

1.1 LA DÉMARCHE DE GREGOR MENDEL

Dans le mémoire publié par Gregor Mendel en 1866, on ne trouve nulle référence aux gènes ou à la génétique, deux mots qui n'apparaissent qu'en 1905. Pas de référence non plus aux chromosomes, qui n'ont pas encore été observés, ni d'ailleurs à l'hérédité. Pourtant, l'intérêt fondateur du travail de Gregor Mendel, est d'avoir abordé l'étude de la variation des caractères entre parents et descendants, en limitant le nombre des caractères, en choisissant un modèle biologique peu sensible aux conditions de l'environnement, et surtout en adoptant une démarche d'analyse quantitative simple, consistant à compter à chaque génération les individus qui présentent le caractère sélectionné par l'expérimentateur. Gregor Mendel mit ainsi en lumière par ces choix, l'existence d'une sorte d'algèbre capable de prédire la distribution des caractéristiques d'un individu en révélant l'existence de causes cachées, conservées dans les cellules sexuelles et transmissibles à la génération suivante selon une combinatoire statistiquement invariable.

Redécouvert au début du XXe siècle après 40 ans d'oubli, le travail de Gregor Mendel suscita immédiatement un engouement extraordinaire. La première tâche fut de tester et de clarifier les trois principes ou « lois » de Mendel : la « loi » de dominance (qu'il n'avait pas



formulée !), la « loi » de disjonction des caractères et la « loi » de ségrégation indépendante des caractères. En raison de nombreuses exceptions, la loi de dominance fut très vite abandonnée, celle de la ségrégation indépendante contestée (il faut que les caractères soient portés par des chromosomes séparés) et celle de disjonction (« loi de pureté des gamètes »), la seule réellement universelle, rencontra certaines exceptions. La plupart de celles-ci s'expliquent par des mécanismes particuliers n'affectant pas fondamentalement la règle. C'est le cas de l'épistasie (voir plus loin) ou altération de l'expression d'un gène par un autre gène ; c'est le cas aussi des allèles multiples, des « polygènes » (différents gènes sont susceptibles de modifier l'expression quantitative d'un même caractère). Tous ces phénomènes ont finalement renforcé le modèle de Mendel, car ils n'affectaient que l'expression des gènes, non les règles de transmission.

Le modèle 3:1

Les lignées parentales sont pures. Croisés entre eux, les individus de chaque lignée transmettent à leurs descendants des caractères homogènes et exclusifs. Les lignées sélectionnées par Gregor Mendel diffèrent chez les parents pour un seul caractère. Il observe et dénombre les deux phénotypes parentaux sur chaque individu de la descendance de première génération (F1, filiation de 1re génération) obtenue par croisement des parents et sur la descendance de deuxième génération (F2) obtenue par croisement des individus F1 entre eux *(Tableau 1.1)*.

Tableau 1.1 Résultats obtenus par Gregor Mendel dans les croisements de lignées pures de pois différant par un caractère.

Phénotypes parentaux en croisement	Phénotype de F1	Phénotype de F2	Rapport des 2 phénotypes de F2
Graines lisses × ridées	100 % lisses	5 474 lisses 1 850 ridées	2,96:1
Graines jaunes × vertes	100 % jaunes	6 022 jaunes 2 001 vertes	3,01:1
Pétales violets × blancs	100 % violets	705 violets 224 blancs	3,15:1
Gousses jaunes × vertes	100 % vertes	428 vertes 152 jaunes	2,82:1
Fleurs axiales × terminales	100 % axiales	651 axiales 207 terminales	3,14:1



Chez les individus F1, l'un des deux **phénotypes** parentaux est dominant et observable, l'autre est récessif et caché. Cependant tous les individus F1 possèdent potentiellement la capacité d'engendrer les deux phénotypes puisque le phénotype récessif réapparaît à la génération F2 et représente le quart des « individus » (rapport des phénotypes F2 proche de 3:1). Par l'examen des individus de la génération F3, obtenue par autofécondation des individus F2, Gregor Mendel révèle que les individus F2 se distribuent en réalité selon un rapport génotypique l:2:l. Autrement dit chaque plante F1 possède une paire de gènes pour un caractère phénotypique : l'un des gènes est responsable du phénotype dominant, l'autre du phénotype récessif. Les deux gènes se distribuent à la méiose dans les gamètes mâle et femelle de sorte que chaque gamète ne possède qu'un des deux gènes parentaux. Au moment de la reproduction, la formation de l'   uf par la fusion des gamètes produit un descendant avec une paire de gènes reconstituée aléatoirement.

Un gène peut exister sous différentes formes dénommées **allèles** (ou allélomorphes, formes alternatives). Un individu, né de la fusion des gamètes (ovule et spermatozoïde), reçoit deux lots de gènes « équivalents » formant un seul ensemble de **gènes appariés** (paires de gènes). On doit dire en réalité paire d'allèles. En effet, chaque gène (nommé par exemple : « *G* ») se trouve représenté chez cet individu sous la forme de deux allèles, l'un maternel, l'autre paternel. Les allèles ainsi réunis peuvent être identiques ou différents. Lorsqu'ils sont identiques l'individu sera dit « **homozygote** » (*GG* ou *gg*) et lorsqu'ils diffèrent, il sera dit « **hétérozygote** » (*Gg*). L'un des deux allèles du gène peut l'emporter sur l'autre dans son expression (l'allèle *G* est dit dominant sur l'allèle *g*, dit récessif). On parlera d'homozygote dominant (*GG*) ou d'homozygote récessif (*gg*) lorsque les 2 allèles sont identiques. Chaque individu possède une formule génétique appelée **génotype** qui regroupe 2 allèles parmi tous les allèles de ce gène présents dans l'espèce. On dira par exemple que *GG* et *Gg* sont deux génotypes distincts, alors que les phénotypes sont généralement similaires puisque *G* domine sur *g*. Le génotype de tout individu possédant deux allèles du gène « *G* » aura au choix la formule *GG* ou *Gg* ou *gg*. On voit ici que gène et allèle sont deux termes associés dans une relation proche de celle qui réunit par exemple les termes genre et espèce. Un genre peut avoir de nombreuses espèces, et une espèce appartient à un genre. Un gène peut avoir de nombreux allèles, et un allèle se rapporte toujours à un gène. À l'échelle moléculaire, la compréhension des deux termes s'avère finalement plus simple. Puisqu'un gène est une séquence en bases de l'ADN, dotée d'une fonction, toute séquence modifiée, même sur une seule base, altérant



ou non la fonction, doit être considérée comme un allèle. Ainsi, pour un gène donné, le nombre d'allèles est théoriquement élevé, mais dans la nature, les allèles n'exerçant plus la fonction correspondante, surtout si elle est vitale, sont en règle générale éliminés avec les individus homozygotes qui les portent. Finalement dans une population formée de nombreux individus, un gène donné pourra se présenter sous la forme de nombreux allèles (cela peut aller jusqu'à plusieurs centaines). Mais chez un individu particulier on ne trouvera au mieux que 2 types d'allèles distincts (individu hétérozygote) ou même un seul type si l'individu est homozygote (les 2 allèles sont identiques).

Le modèle 9:3:3:1

Les lignées parentales sont pures (homozygotes) et diffèrent par deux gènes contrôlant deux caractères séparés (4 phénotypes). Comme précédemment, on dénombre les individus des générations F1 obtenus par croisement des parents et F2 par autofécondation entre individus F1.

Des calculs effectués pour chacun des couples de caractères, on peut conclure qu'ils sont indépendants puisque l'un et l'autre répondent au modèle 3:1. Le rapport 9:3:3:1 n'est finalement que le produit aléatoire de deux rapports indépendants 3:1 *(Tableau 1.2)*.

Tableau 1.2 Résultats obtenus par Gregor Mendel dans le croisement de deux lignées pures de pois différant par deux caractères.

Phénotypes parentaux	Phénotype de F1	Phénotype de F2	Rapport des phénotypes de F2	Rapport des phénotypes de F2 pour le caractère lisse/ridé	Rapport des phénotypes de F2 pour le caractère jaune/vert
Graines lisses et vertes × Graines ridées et jaunes	100 % de graines lisses et jaunes	315 lisses et jaunes	9	$\frac{423}{133}$ (3,18:1)	$\frac{416}{140}$ (2,97:1)
		108 lisses et vertes	3		
		101 ridées et jaunes	3		
		32 ridées et vertes	1		



Vérification expérimentale des deux modèles par le croisement test

Vérification du modèle 3:1

Le croisement de pois à graines jaunes (*JJ*) et de pois à graines vertes (*jj*) donne *(Tableau 1.1)* une génération F1 à phénotype dominant, à graines jaunes et de génotype hétérozygote (*Jj*). En croisant un hybride F1 (*Jj*) avec une lignée homozygote récessive à graines vertes (*jj*) on peut prédire *(Fig.1.1)* qu'il y aura autant de graines jaunes que de graines vertes dans la descendance. Le **croisement test** prédit donc un rapport de phénotypes égal à 1:1.

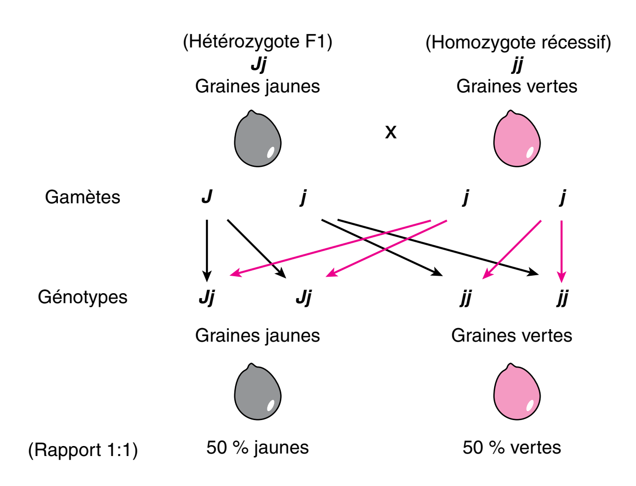


Figure 1.1 Croisement d'un pois hétérozygote (hybride F1) avec un pois homozygote récessif.

Vérification du modèle 9:3:3:1

Le croisement de pois à graines lisses et vertes avec des pois à graines ridées et jaunes *(Tableau 1.2)* donne une génération F1 où toutes les graines sont lisses et jaunes, avec un génotype sous-jacent hybride *RrJj*. En croisant un tel individu F1 double hétérozygote avec une lignée double homozygote récessif *(Fig. 1.2)* on peut prédire qu'il y aura équiprobabilité de rencontrer les 4 phénotypes (graines jaunes et ridées, vertes et ridées, lisses et jaunes et lisses et vertes) dans la descendance, soit un rapport 1:1:1:1.

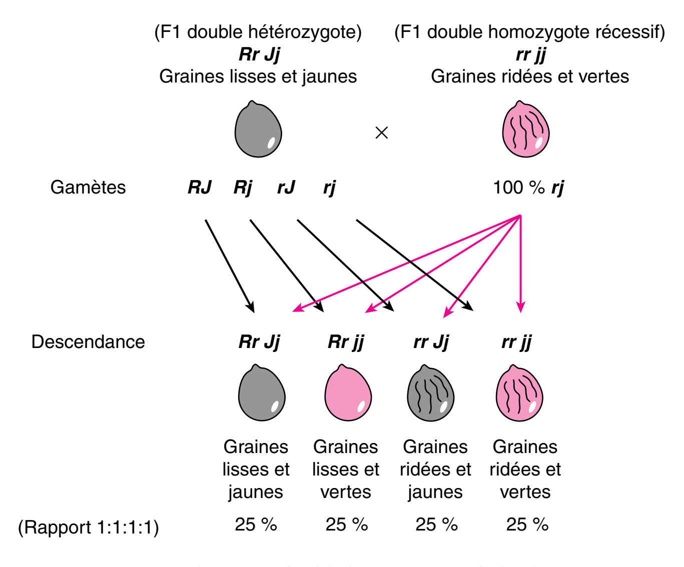


Figure 1.2 Croisement d'un pois, double hétérozygote (hybride F1), avec un pois double homozygote récessif?

Le test de croisement s'opère donc entre l'hétérozygote de la génération F1 et un homozygote récessif. Il permet à l'expérimentateur d'analyser le génotype sous-jacent à un phénotype dominant puisque le parent homozygote récessif n'apporte que les allèles récessifs à sa descendance. Il peut être appliqué d'une manière générale à la détermination du génotype d'un individu.

Méthode de calcul des rapports génétiques dans les hybrides

Elle est basée sur le produit des rapports et s'applique aux phénotypes comme aux génotypes.

Diagramme pour le double hybride

Lorsqu'on écrit (*R*-) ou (*J*-) cela signifie *(Fig. 1.3)* que l'allèle dominant (*R* ou *J*) qui impose le phénotype peut être indifféremment apparié avec un autre allèle dominant (*RR* ou *JJ*) ou avec un allèle récessif (*Rr* ou *Jj*) sans modification du phénotype.



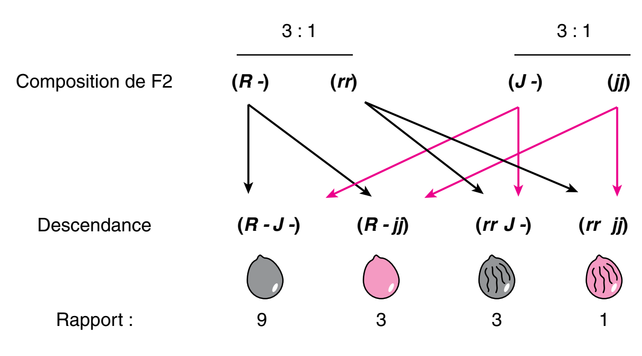


Figure 1.3 Produits des rapports génétiques pour un double hybride.

Généralisation : exemple d'un diagramme pour un triple hybride

On admet comme précédemment qu'il y a ségrégation indépendante des trois gènes correspondants aux trois couples de caractères avec dominance d'un des deux allèles. À la génération F2, on a la descendance correspondant au croisement ($AaBbCc \times AaBbCc$). Le diagramme s'écrit *(Fig. 1.4)* :

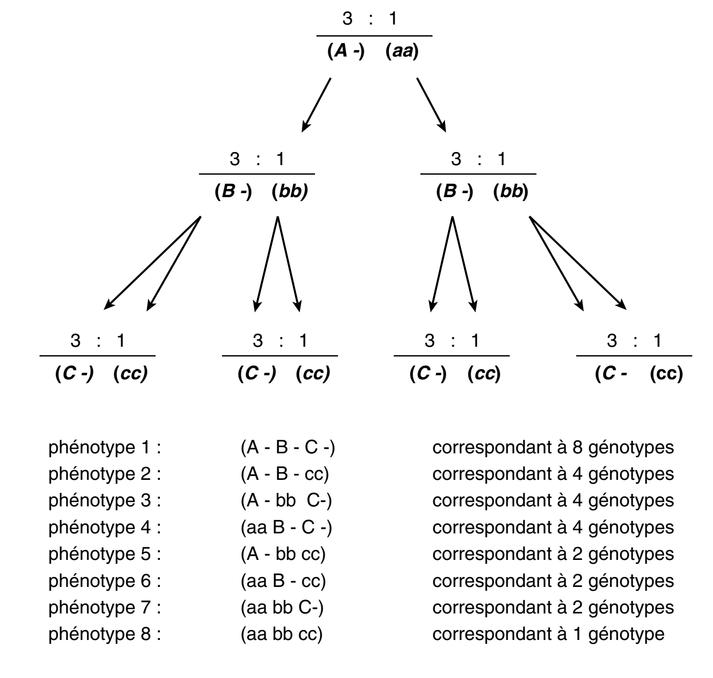


Figure 1.4 Généralisation du calcul des rapports génétiques dans les hybrides. Exemple d'un triple hybride.



On obtient, en effectuant les produits successifs (suivre les flèches de haut en bas), le modèle 27:9:9:9:3:3:3:1 qui n'est qu'une variante combinée du modèle 3:1, avec 8 phénotypes et 27 génotypes *(Fig. 1.4)*.

Avec l'accroissement de gènes appariés participant aux croisements hybrides, tout en maintenant le principe de ségrégation indépendante, le nombre de phénotypes et de génotypes devient rapidement très élevé *(Tableau 1.3)*.

Tableau 1.3 Nombre de phénotypes et de génotypes en fonction du nombre de gènes apparentés dans les croisements hybrides

Nombre de gènes appariés	Nombre de phénotypes	Nombre de génotypes
1	2	3
2	4	9
3	8	27
4	16	81
–	–	–
n	2^n	3^n

1.2 LE MODÈLE 3:1 CHEZ L'HOMME

De nombreux caractères phénotypiques déterminés par des paires d'allèles chez l'homme sont transmis suivant les mêmes principes que ceux qui ont été découverts par Mendel chez le pois. Mais l'analyse génétique chez l'homme doit évidemment respecter les droits de la personne humaine ce qui exclue toute expérimentation, c'est pourquoi le généticien examine les populations humaines naturelles afin d'y rechercher les caractères phénotypiques susceptibles d'être analysés à travers plusieurs générations. Les plus remarquables sont les anomalies récessives qui apparaissent dans la descendance de personnes ne présentant pas elles-mêmes le phénotype anormal. Une fois l'anomalie identifiée, sa transmission est reconstituée dans l'arbre généalogique de la famille où elle est apparue en remontant le plus loin possible dans le temps. Un exemple bien connu est fourni par une maladie humaine bien étudiée et maîtrisée : la phénylcétonurie.

La phénylcétonurie

La phénylalanine hydroxylase, première enzyme de la voie catabolique conduisant de la phénylalanine au fumarate et à l'acétoacétate, catalyse l'hydroxylation de la phénylalanine en tyrosine *(Fig. 1.5)*. Un



déficit en cette enzyme, dû à un défaut du gène, s'avère responsable de la phénylcétonurie, une maladie génétique humaine assez fréquente (8 cas pour 100 000 naissances). Elle est due au fait qu'une voie secondaire du métabolisme de la phénylalanine est activée en raison du déficit en phénylalanine hydroxylase. L'acide aminé subit une transamination et donne le phénylpyruvate *(Fig. 1.5)*, un composé qui s'accumule dans le sang et les tissus et passe avec la phénylalanine dans l'urine, d'où le nom donné à la maladie.

NH_3^+
$-CH_2-CH-COO^-$
Phénylalanine
O_2
H_2O
NADH + H^+
NAD^+
Phénylalanine hydroxylase
HO
NH_3^+
$-CH_2-CH-COO^-$
Tyrosine

Phénylacétate
NH_3^+
$-CH_2-CH-COO^-$
Phénylalanine
$CH_3-CO-COO^-$
Pyruvate
O
$-CH_2-C-COO^-$
Phénylpyruvate
Phényllactate
NH_3^+
$CH_3-C-COO^-$
Alanine

Figure 1.5 Voie normale du catabolisme de la phénylalanine en tyrosine et voie secondaire conduisant au phénylacétate en l'absence d'une phénylalanine hydroxylase active.

Le phénylpyruvate peut être décarboxylé en phénylacétate qui donne une odeur caractéristique à l'urine, un signe qui fut jadis utilisé par les soignants pour détecter la maladie chez le jeune enfant. L'accumulation de la phénylalanine et de ses dérivés au début de la vie perturbe le transport et la disponibilité des autres acides aminés ce qui altère le développement normal du cerveau et provoque un retard mental sévère. La phénylcétonurie fut parmi les premières anomalies génétiques du métabolisme, découvertes chez l'homme. Quand elle est reconnue suffisamment tôt, le retard mental peut être évité par un régime alimentaire strict. La consommation d'aliments riches en protéines est proscrite comme celle, par exemple, d'aliments édulcorés par l'aspartame qui est un dipeptide contenant de la phénylalanine.



Afin de fournir à l'enfant un aliment correct, la caséine du lait est hydrolysée en acides aminés, et la plus grande partie de la phénylalanine en est retirée. Certains biotechnologues envisagent aujourd'hui de créer des vaches transgéniques dont les protéines du lait seraient constitutivement à faible teneur en cet acide aminé.

Depuis quelques dizaines d'années, des tests biochimiques (test de Guthrie) peu coûteux, sont pratiqués dès la naissance, à partir d'une goutte de sang afin de détecter cette maladie, ce qui évite bien des soins coûteux et beaucoup de détresse.

Les enfants atteints par cette maladie (phénotype récessif) possèdent le génotype *pp*, les autres le génotype *PP* ou *Pp*. Généralement cette maladie apparaît chez des enfants (fille et garçon, il n'y a pas de liaison au sexe) de parents sains *(Fig. 1.6* et *Fig. 1.7)*. En effet, puisque seuls les enfants porteurs d'allèles récessifs sont atteints, on en déduit que les parents sont hétérozygotes sains et possèdent chacun un allèle récessif. Les génotypes de l'arbre généalogique familial s'écrivent comme dans la *figure 1.6*.

En raison généralement du trop petit nombre d'individus composant la famille au regard des croisements expérimentaux indispensables pour obtenir le rapport génétique mendélien 3:1 caractéristique du monohybridisme, il est très rare que le rapport observé chez l'homme soit correct (dans la *figure 1.6*, il est de 1:1). Pour connaître la fréquence relative des 3 génotypes de cette maladie dans une population humaine en équilibre, il est possible d'utiliser le calcul suivant : si la proportion relative des deux allèles *P* et *p* est x et y respectivement, la fréquence des 3 génotypes possibles sera de x^2 pour *PP*, $2\,xy$ pour *Pp* et y^2 pour *pp*. Numériquement pour une fréquence d'homozygotes potentiellement malades de $y^2 = 1/12\,500$, soit 8 %, la fréquence y de l'allèle récessif *p* sera environ de l/l 12, la fréquence x de l'allèle dominant *P* de 111/112 et la fréquence des hétérozygotes *Pp* de $2\,xy = 2\,(1/112)\ (111/l12) = 1/56$. Ce qui signifie que les porteurs sains (*Pp*) sont 220 fois plus fréquents que les malades potentiels (homozygotes récessifs, *pp*), une différence qui ne peut aller

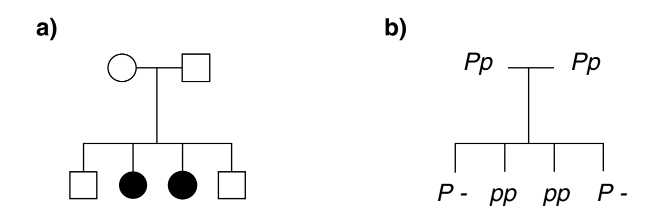


Figure 1.6 Descendance de parents sains porteurs hétérozygotes d'une anomalie génétique. **a)** Cas général ; **b)** application à la phénylcétonurie.



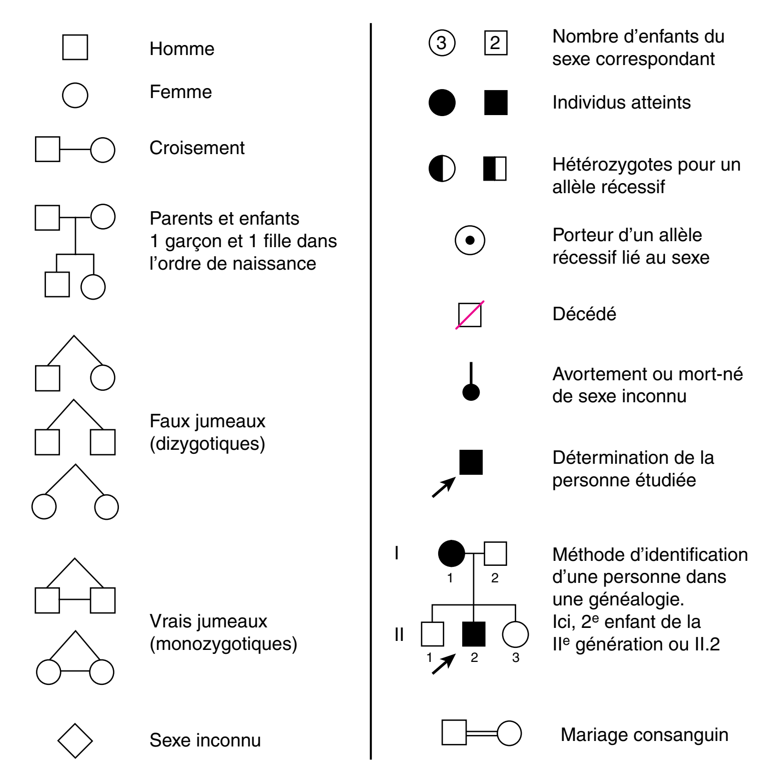


Figure 1.7 Symboles utilisés dans l'analyse généalogique.

qu'en augmentant dans les générations successives et qui souligne finalement la rareté de l'allèle *p*. À l'inverse, les unions consanguines accroissent le risque de naissance d'un enfant malade particulièrement si les ancêtres des parents sont hétérozygotes *(Fig. 1.8)*. Cet exemple, qui peut être généralisé (la mucoviscidose ou l'albinisme sont deux autres exemples bien documentés), illustre pourquoi les mariages entre personnes apparentées, cousins par exemple, augmentent la fréquence des maladies récessives dans les populations humaines.

Cependant, grâce aux progrès de la génétique moléculaire et des techniques d'identification des gènes et de leurs mutations, la prédiction du risque d'anomalies génétiques liées notamment aux allèles récessifs s'avère de plus en plus précise et précoce. Il existe chez l'homme également des maladies génétiques transmises selon un mode dominant (par exemple la chorée de Huntington). Dans ces cas, l'allèle normal est récessif et c'est l'allèle anormal, muté, qui est dominant. La maladie (ou le désordre génétique) peut alors apparaître à chaque génération et il est de ce fait en théorie possible de prévoir et d'éviter sa transmission.



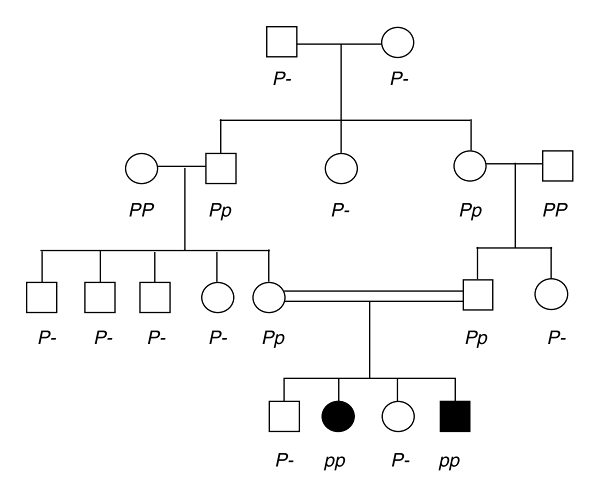


Figure 1.8 Généalogie d'une famille où deux enfants sont atteints de phénylcétonurie (homozygotes récessifs *pp*) à la suite du mariage consanguin entre des cousins dont les ascendants sont hétérozygotes.

De tels allèles dominants défectueux peuvent également résulter de mutations spontanées (par exemple au moment de la formation des gamètes) ou provoquées par des conditions environnementales particulières (par exemple : à la suite d'une irradiation radioactive ou d'une exposition à un agent mutagène). La personne concernée peut ne pas en subir immédiatement les conséquences sauf en général si l'événement se produit à un stade embryonnaire précoce. Mais elle peut transmettre la mutation à sa descendance si sa lignée germinale (gamètes) est atteinte.

La perception de l'amertume par l'homme et les grands singes

Il est banal de constater de discrètes différences morphologiques entre deux individus de la même espèce, même très proches génétiquement. Elles ne peuvent pas être pour autant attribuées à des anomalies génétiques. Dans l'espèce humaine, elles sont à la base des processus de reconnaissance entre les individus. Dans toutes les espèces, ces phénotypes normaux, mais différents, constituent ce que le généticien appelle le polymorphisme. Dans une version simplifiée, les allèles de plusieurs gènes portés notamment par les chromosomes X et Y sont impliqués. On parle de dimorphisme entre les individus de sexe mâle ou femelle (dimorphisme sexuel).

Le polymorphisme peut aussi se présenter sous une forme cachée. C'est le cas par exemple dans la capacité des humains et des grands



singes à percevoir le goût amer d'une substance chimique de synthèse, le phénylthiocarbamide (PTC) qui contient le groupe chimique (N–C=S). Certains individus perçoivent une amertume très désagréable, alors que d'autres ne perçoivent rien du tout. La région génomique, en cause chez l'homme, contient le gène *TAS2R38* qui code l'un des multiples récepteurs du goût amer. Cinq allèles (haplotypes) résultant de la présence de 3 mutations ponctuelles dans la région codante produisent différents récepteurs qui expliquent complètement la distribution bimodale de la sensibilité ou non au PTC. Conduit au sein de différentes familles, un test sensoriel montre, sur la base d'une transmission de type mendélien, que deux parents sensibles peuvent avoir des descendants insensibles *(Fig. 1.9)*. Il en résulte que l'allèle principal (l'**haplotype**) de sensibilité est dominant, et ceux de l'insensibilité, récessifs. Bien qu'aucun des deux phénotypes n'ait une valeur sélective clairement identifiée, leur apparition chez l'homme et le chimpanzé résulterait d'événements mutationnels totalement indépendants.

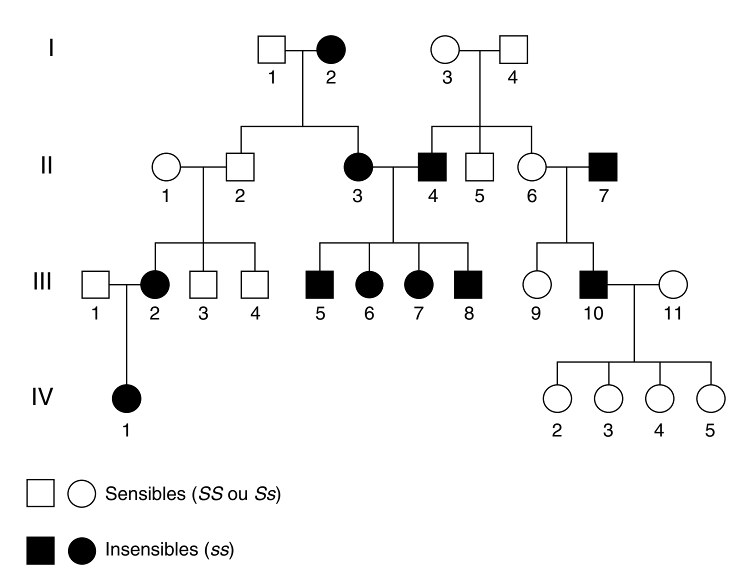


Figure 1.9 Généalogie d'un dimorphisme humain : l'aptitude à percevoir ou non l'amertume du phénylthiocarbamide (PTC).

1.3 DOMINANCE ET RÉCESSIVITÉ, MULTIALLÉLISME

Aspects généraux

La notion de **dominance** d'un allèle s'oppose à celle de **récessivité**. Elle résulte directement de l'état diploïde des métazoaires (individus pluricellulaires engendrés au sein d'une espèce par reproduction sexuée).



Ces organismes possèdent dans chacune des cellules de leur corps une copie de chaque chromosome en provenance de l'un et l'autre des parents. Les deux chromosomes parentaux, dits homologues, portent ainsi la même série de gènes. Pour un gène donné, l'allèle paternel peut différer cependant de l'allèle maternel, et inversement. L'un des allèles est dit muté, comparativement à l'autre. Si les deux allèles portés par un individu sont identiques, le phénotype qu'ils influencent sera en général exprimé. S'ils sont différents, le phénotype de l'individu, dit alors hétérozygote, dépend de l'allèle qui est dominant. L'autre allèle, dit récessif, peut correspondre à un phénotype particulier qui ne sera pas exprimé. On dit qu'il est réprimé. Il s'agit d'un cas de simple dominance. Tout au long de la descendance de l'individu hétérozygote, le trait phénotypique qui dépend de l'allèle dominant sera exprimé à chaque génération, alors que ce n'est pas le cas pour le trait phénotypique qui dépend de l'allèle récessif. Un descendant ne peut exprimer le phénotype « récessif » que, si les deux allèles sont récessifs et identiques (état homozygote).

La dominance de l'allèle sera dite incomplète si le phénotype de référence, dit de type sauvage (parce que le plus répandu dans la population naturelle) dépend de la présence des deux allèles, dits aussi de type sauvage. La présence d'un seul allèle sauvage peut conduire à un phénotype intermédiaire, d'où le terme de **dominance incomplète**. Il arrive parfois que l'allèle muté, généralement récessif, influence cependant l'expression du phénotype dépendant de l'autre allèle. Cela peut se produire par exemple si la protéine produite par l'allèle muté inhibe par compétition la protéine produite par l'allèle sauvage. On parle alors de **dominance négative**. Ce phénomène s'observe très souvent lorsqu'un phénotype dépend d'un complexe protéique fonctionnel composé de plusieurs sous-unités (enzymes ou récepteurs de voies de signalisation, voir ci-dessous)

Le terme de **co-dominance** désigne la situation où les deux allèles de type sauvage ou non expriment simultanément et avec une « force » comparable deux phénotypes distincts (voir également ci-dessous).

Illustration par un exemple simple

Supposons qu'un allèle dit de type sauvage puisse coder au minimum la synthèse de 1500 molécules d'une enzyme monomérique (c'est-à-dire constituée d'une seule chaîne polypeptidique) responsable d'un trait phénotypique très majoritaire au sein d'une population étudiée. Les deux allèles de ce type coderont ensemble au minimum 3 000 molécules d'enzymes. Si l'un des deux allèles est porteur d'une mutation qui conduit à une enzyme non fonctionnelle, l'autre allèle peut coder suffisamment de molécules d'enzyme fonctionnelle (1500)



pour que le phénotype sauvage continue à être pleinement exprimé, si par exemple seules 1000 molécules d'enzymes sont nécessaires pour cela. On dira de cet allèle de type sauvage qu'il est dominant. Il peut arriver que l'allèle muté soit responsable d'une activité enzymatique distincte de celle qui est associée à l'allèle de type sauvage. En présence des deux allèles, un phénotype intermédiaire peut en résulter qui marquera l'existence d'une co-dominance des deux allèles. Par contre si les deux allèles mutés sont identiques et ne codent pas d'enzyme fonctionnelle, le phénotype sauvage ne sera plus exprimé. Si cette absence d'enzyme n'est pas létale, c'est le phénotype récessif qui prévaudra.

Étudions maintenant le cas d'une enzyme dimérique fonctionnelle, constituée de deux sous-unités polypeptidiques identiques. Les deux allèles de type sauvage coderont la formation de 1500 molécules d'enzyme dimérique et le phénotype correspondant sera comme précédemment de type sauvage. Mais si l'un des deux allèles est muté, l'autre allèle de type sauvage ne pourra plus coder que 375 molécules d'enzyme dimérique fonctionnelle (une seule configuration parmi les 4 possibles, 1500:4). Ce niveau d'enzyme peut s'avérer insuffisant pour que le phénotype sauvage s'exprime pleinement si comme précédemment 1000 molécules sont nécessaires pour cela. On parlera alors de dominance incomplète.

Il peut arriver cependant que l'allèle sauvage produise suffisamment de sous-unités de l'enzyme (par exemple 4000) pour que la quantité finale de celle-ci, à l'état dimérique (dans cet exemple 4000:4, soit 1000) soit suffisante pour que le phénotype sauvage s'exprime pleinement.

Supposons pour continuer la démonstration que l'enzyme soit tétramérique (constituée de quatre chaînes polypeptidiques identiques). Les deux allèles sauvages doivent produire chacun au minimum 2000 molécules d'enzymes pour qu'au final, 1000 molécules d'enzyme tétramérique puissent se former et produire un phénotype sauvage complet dans les conditions définies plus haut. Si comme dans le premier cas évoqué, chacun des allèles ne produit que 1500 sous-unités enzymatiques, soit 3000 sous-unités au total, seules 750 molécules d'enzyme tétramérique fonctionnelle pourront se former et le phénotype correspondant sera de type intermédiaire. Si l'un des deux allèles est porteur d'une mutation conduisant à une sous-unité non fonctionnelle, alors seules 1500 sous-unités seront produites par l'allèle de type sauvage et un seul tétramère fonctionnel, sur les 12 configurations possibles, se formera soit 125 tétramères enzymatiques fonctionnels (1500:12) ce qui est probablement très insuffisant pour qu'un phénotype puisse en résulter. On considère dans ce cas que la (les) mutation(s)



de l'allèle inactivant la sous-unité enzymatique sont de type dominant négatif.

Le cas des groupes sanguins ABH

Il arrive fréquemment qu'un gène au sein d'une population possède un nombre élevé d'allèles différents. Un individu diploïde de cette population n'est par définition porteur que de deux exemplaires du gène, c'est-à-dire deux allèles seulement (identiques ou différents). C'est l'examen de différents individus de la même population qui révèle en général ce **multiallélisme**. On dit que la collection des différents allèles constitue une série allélique. Un exemple simple en est donné par les allèles des groupes sanguins humains ABH (0).

Dans le système ABH, on connaît quatre phénotypes *(Tableau 1.4)* et trois allèles principaux (I^B, I^A et i). Chaque être humain possède seulement l (en double exemplaire) ou 2 allèles parmi les 3 possibles. Les allèles I^A et I^B déterminent la synthèse à la surface des cellules (hématies) d'un antigène unique. L'allèle i correspond à l'absence de synthèse de l'un des 2 antigènes précédents. Les allèles I^A et I^B sont dominants lorsqu'ils sont associés à i, et codominants dans le génotype I^AI^B.

TABLEAU 1.4 GROUPES SANGUINS ABH HUMAINS.

Groupe (phénotype)	Génotype
H (O, ni A, ni B)	ii
A	I^AI^A ou I^Ai
B	I^BI^B ou I^Bi
AB	I^AI^B

Le cas du gène de la tyrosinase

Un autre exemple est fourni par le gène *C* des mammifères. Il code une enzyme, la tyrosinase, qui intervient dans la synthèse des pigments (mélanines) de la robe. Les différents allèles produisent des phénotypes variables allant du noir (synthèse pleinement active) à l'absence de pigments (albinisme) en passant par la série des gris (synthèse plus ou moins active). Les relations de dominance/récessivité entre les allèles d'une série sont parfois complexes, tel allèle pouvant être à la fois dominant et récessif suivant l'autre allèle avec lequel il est combiné chez un individu diploïde. L'allèle le plus fréquemment rencontré dans une population naturelle sera généralement



appelé allèle de type sauvage. Chez les végétaux, le multiallélisme est parfois impressionnant pouvant aller pour certains gènes particuliers, jusqu'à plusieurs centaines d'allèles.

Dénommer les allèles n'est donc pas toujours simple, de sorte que les symboles employés par les différents laboratoires peuvent varier. Des efforts d'harmonisation de la nomenclature sont cependant accomplis par les différentes communautés scientifiques regroupées dans l'étude de quelques organismes modèles.

1.4 ALLÈLES LÉTAUX

Un exemple de ce type d'allèle a été analysé par Lucien Cuenot, un généticien français du début du XXe siècle. Des souris à pelage clair (jaunâtre) naissent parfois de parents à pelage de type sauvage (on parle d'un phénotype « agouti » le poil sauvage étant de couleur noire avec une bande jaune, due à l'expression transitoire du gène *A* (pour « Agouti »). Croisés entre eux, les individus F1 à pelage clair donnent une descendance composée toujours aux 2/3 d'individus « jaunes » et pour 1/3 d'individus de type sauvage, alors que le rapport mendélien attendu serait de 1:2:1.

Lucien Cuenot a proposé qu'un allèle A^Y, responsable de la couleur claire (y pour yellow : jaune) soit dominant sur l'allèle sauvage *A*. À l'état homozygote, l'allèle A^Y serait de plus létal (A^YA^Y). Les individus de ce type ne seront donc pas comptabilisés dans la descendance du fait de leur mort précoce *in utero*. Ainsi, dans le croisement des F1 ($AA^Y \times A^YA$) le rapport devient bien 2/3 A^YA (« jaune ») et 1/3 *AA* (« sauvage »), après élimination des embryons A^YA^Y, c'est-à-dire le rapport observé. De plus, des souris de phénotype sauvage (génotype *AA*) croisées avec des souris de phénotype « jaune » (A^YA) donneront toujours une descendance à 50 % de phénotype sauvage et 50 % de phénotype « jaune ».

On dit d'un gène qui induit des effets distincts (ici pelage clair et viabilité) qu'il est **pléiotrope**. Le gène Agouti murin possède d'autres allèles qui confirment ce phénomène. Par exemple, l'allèle A^{VY} (viable et jaune) donne un pelage clair et des souris obèses. Il n'a pas encore été possible de découvrir à l'échelle moléculaire comment la protéine Agouti pouvait provoquer ces effets aussi distincts. On doit remarquer cependant que les phénotypes peuvent dépendre de la concentration de la protéine puisque le gène A^Y par exemple est létal à l'état homozygote (double dose) mais pas à l'état hétérozygote (simple dose). Plus récemment on a découvert que l'expression du gène Agouti est contrôlée par des mécanismes épigénétiques liés à des niveaux variables de méthylation de certaines régions de l'ADN, régulatrices de la transcription du gène (voir chapitre 5).



1.5 ACTUALITÉ DU CONCEPT DE GÈNE

Un concept historique

Dans les années 1860, Gregor Mendel suggère pour la première fois (voir plus haut), l'existence de « facteurs » transmissibles des parents à leurs descendants. Bien que le terme même de **gène** soit inventé plus tard, il explique ses résultats en termes d'hérédité des caractères, fait l'hypothèse de leur assortiment indépendant, distingue les traits **dominants** et **récessifs**, formule l'existence d'**hétérozygote** et d'**homozygote**, avance ce qui ultérieurement sera défini comme le **génotype** et le **phénotype**. En 1889, le terme de « pangène », abrégé vingt ans plus tard en « gène », est inventé par Hugo de Vries qui, tout en ignorant les travaux de Mendel, désigne par ce mot la plus petite particule correspondant à une caractéristique héréditaire. En 1910, Thomas Morgan montre que les gènes sont portés par les chromosomes, où ils occupent des positions spécifiques ; il crée avec ses élèves la première carte chromosomique de la drosophile. En 1928, Frederick Griffith démontre que les gènes peuvent se transmettre : des bactéries pathogènes pour la souris, bien que tuées par la chaleur, peuvent transmettre une information génétique et transformer d'autres bactéries, vivantes mais non pathogènes, au point de les rendre capables de tuer à leur tour des souris. En 1944, Oswald Theodore Avery, Colin McLeod et Maclyn McCarty montrent que le principe chimique transformant les bactéries est l'ADN (Acide DésoxyriboNucléique), et non les protéines comme cela était l'hypothèse dominante. Ils établissent que l'ADN est le support de l'hérédité. En 1953, en se basant sur les travaux de cristallographie de Rosalind Franklin, James D. Watson et Francis Crick établissent la structure en double hélice de l'ADN. Cette découverte majeure apporte pour la première fois une explication strictement physico-chimique à la notion d'information génétique et à sa transmission de génération en génération. Dans une phrase devenue célèbre, les deux auteurs indiquent en effet, à la fin de leur article fondateur : « *Il ne nous a pas échappé que l'appariement spécifique postulé suggère du même coup, un possible mécanisme de reproduction du matériel génétique*. » Ils anticipaient, ce qui fut ultérieurement démontré, que chaque brin de l'ADN après séparation, servait de matrice pour la synthèse d'un brin complémentaire, et qu'à partir d'une molécule d'ADN on pouvait en obtenir deux, ce qui est la marque même du principe de la reproduction et de la transmission physique, *via* un support matériel, des instructions génétiques caractéristiques de chaque être vivant.

Pour la plupart des biologistes, la notion de gène, qui était alors considérée comme un concept plutôt abstrait, devint du coup une réalité physique isolable et manipulable. Le gène se définit alors



comme une unité chimique héréditaire, codée dans une séquence de nucléotides, ces derniers étant eux-mêmes, organisés en de longs brins d'ADN, constitutifs des chromosomes. Un gène ainsi défini, peut adopter de multiples formes différentes (les allèles du gène) correspondant à différentes séquences nucléotidiques très proches les unes des autres, et donc apparentées. Après avoir été converties en ARN messagers, ces séquences d'ADN sont à l'origine de la construction des protéines correspondantes, dans le respect du principe qu'à un gène correspond au moins une protéine.

Pour de nombreux biologistes, cette vision devenue historique et conventionnelle reste acceptable, à quelques détails près. Pour un nombre grandissant de généticiens, il ne s'agit là que d'une approximation cachant une réalité bien plus complexe.

Un concept dynamique

Au cours des dernières décades en effet, des découvertes sur la manière dont les informations génétiques associées à la séquence de l'ADN assurent le fonctionnement des êtres vivants ont révolutionné en la complexifiant la vision traditionnelle du concept de gène. Contrairement au fonctionnement des gènes procaryotes, à l'origine du principe « un gène, une protéine », on a découvert, en premier lieu, que la séquence codante des gènes eucaryotes est discontinue, en quelque sorte dispersée au sein de régions non codantes. Le concept de **gène morcelé** d'une part en **exons** (parties codantes) généralement de petite taille et d'autre part en **introns** de plus grande taille (parties non codantes) en résulta. Pour reconstituer un ARN messager, apte à diriger la synthèse des protéines, les transcrits primaires d'un tel gène sont épissés, grâce à un mécanisme qui élimine les introns et reconstitue une séquence codante continue. Les nombreuses variantes d'un tel mécanisme aboutissent à différents ARN messagers alternatifs et finalement à la synthèse de plusieurs protéines différentes. Contrairement à la vision classique un seul gène peut donc engendrer plusieurs protéines différentes grâce à un épissage alternatif (voir le Mini Manuel *Biologie moléculaire* dans cette collection).

Un degré de complexité supplémentaire fut révélé avec l'existence de gènes dont les séquences sont chevauchantes (des exons et introns sont communs à plusieurs gènes et d'autres non). Autres éléments plus récents de complexification, des gènes sont placés parfois à l'intérieur d'introns d'autres gènes ; des transcrits provenant de régions génomiques distinctes se recombinent en une seule molécule d'ARN préalablement à leur traduction. Le développement des connaissances portant sur des génomes entiers révèle que de très



nombreux transcrits sont synthétisés sans qu'ils soient ensuite traduits en protéines. Il s'agit d'ARN « messagers » non codants. Correspondent-ils à des gènes au sens traditionnel du mot ? Ces découvertes attestent qu'un grand nombre d'informations, peut-être majoritaires, et apparemment indispensables à la formation et au fonctionnement des êtres vivants, transitent par des transcrits ARN et non par des protéines *(Fig. 1.10)*. Plus étonnant encore, certains ARN associés à des phénotypes clairement identifiables, semblent être transmis de génération en génération, indépendamment des « gènes » dont ils sont issus (voir chapitre 2, la notion de paramutation).

Aussi, à la question « Qu'est-ce qu'un gène ? », la réponse à ce jour s'avère très incertaine. On pourrait dire d'une façon certes un peu vague, qu'il s'agit « d'une région définie de la molécule d'ADN, correspondant à une unité héréditaire, localisable au sein d'une séquence génomique, et qui se trouve associée à des régions régulatrices, des régions transcrites et d'autres régions fonctionnelles ».

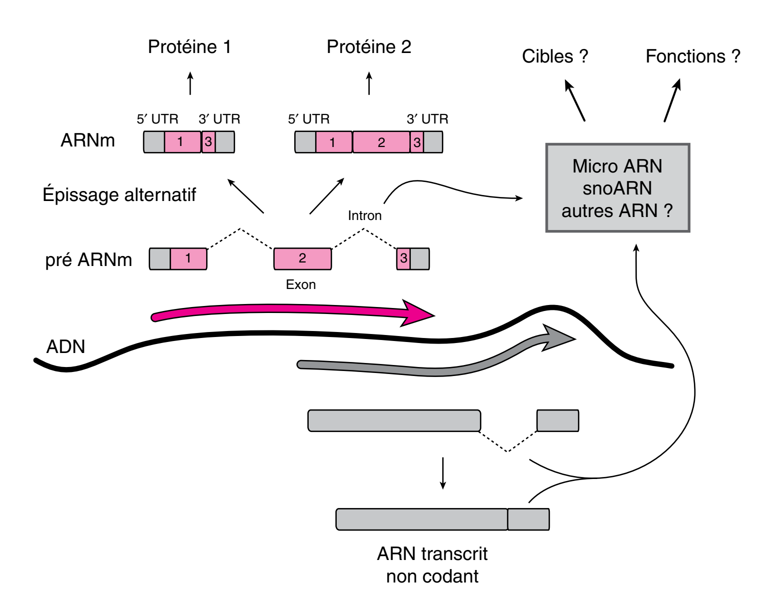


Figure 1.10 Schéma illustrant la complexité d'une région génomique contenant des séquences codant des protéines et des séquences produisant des ARN non codants.

Les transcrits dérivent de l'un ou l'autre des deux brins d'ADN. Ils sont parfois chevauchants ou recombinants. Certains, y compris les transcrits non codants sont alternativement épissés. À la fois les exons et les introns peuvent transmettre des informations génétiques. De nombreux microARNs et petits ARN nucléolaires dérivent des introns chez les animaux. Leurs typologies et leurs fonctions sont encore largement inconnues.



Malgré le flou d'une telle définition, il n'en demeure pas moins que le concept (la notion ou l'objet) sous-jacent au terme de « gène » reste d'un emploi commode pour désigner les entités génétiques transmises de génération en génération.



POINTS CLEFS

- L'intérêt fondateur du travail de G. Mendel, est d'avoir abordé l'étude de la variation des caractères entre parents et descendants, en limitant le nombre des caractères, en choisissant un modèle biologique peu sensible aux conditions de l'environnement, et surtout en adoptant une démarche d'analyse quantitative simple, consistant à compter à chaque génération les individus qui présentent le caractère sélectionné par l'expérimentateur.
- Les trois principes ou « lois » de Mendel sont : la « loi » de dominance (qu'il n'avait pas formulée !), la « loi » de disjonction des caractères et la « loi » de ségrégation indépendante des caractères. En raison de nombreuses exceptions, la loi de dominance fut très vite abandonnée, celle de la ségrégation indépendante contestée (il faut que les caractères soient portés par des chromosomes séparés) et celle de disjonction (« loi de pureté des gamètes »), la seule réellement universelle, rencontra certaines exceptions. Le modèle de Mendel reste une référence, car la plupart des exceptions n'affectent que l'expression des gènes, non les règles de transmission.
- Un gène peut exister sous différentes formes dénommées allèles (ou allélomorphes, formes alternatives). Un individu, né de la fusion des gamètes (ovule et spermatozoïde), reçoit deux lots de gènes « équivalents » l'un maternel, l'autre paternel, formant un seul ensemble de gènes appariés (paires de gènes).
- Gène et allèle sont deux termes associés dans une relation proche de celle qui réunit par exemple les termes genre et espèce. Un gène peut avoir de nombreux allèles, et un allèle se rapporte toujours à un gène.
- Dans une population formée de nombreux individus, un gène donné pourra se présenter sous la forme de nombreux allèles (cela peut aller jusqu'à plusieurs centaines). Mais chez un individu particulier on ne trouvera au mieux que 2 allèles distincts (individu hétérozygote) ou même un seul si l'individu est homozygote (les 2 allèles sont identiques).
- La notion de dominance d'un allèle s'oppose à celle de récessivité. Elle résulte directement de l'état diploïde des métazoaires (individus pluricellulaires engendrés au sein d'une espèce par reproduction sexuée). Ces organismes possèdent dans chacune des cellules de leur corps une copie de chaque chromosome en provenance de l'un et l'autre des parents.



- Les relations de dominance/récessivité entre les allèles d'une série sont parfois complexes, tel allèle pouvant être à la fois dominant et récessif suivant l'autre allèle avec lequel il est combiné chez un individu diploïde. L'allèle le plus fréquemment rencontré dans une population naturelle sera généralement appelé allèle de type sauvage.
- On dit d'un gène qui induit des effets distincts qu'il est pléiotrope.
- La structure en double hélice de l'ADN est une découverte majeure qui a apporté pour la première fois une explication strictement physico-chimique à la notion d'information génétique et à sa transmission de génération en génération. À partir d'une molécule d'ADN on peut en obtenir deux, ce qui est la marque même du principe de la reproduction et de la transmission physique, *via* un support matériel, des instructions génétiques caractéristiques de chaque être vivant.
- À la question « Qu'est-ce qu'un gène ? », la réponse s'avère encore aujourd'hui très incertaine. On peut avancer qu'il s'agit « d'une région définie de la molécule d'ADN, correspondant à une unité héréditaire, localisable au sein d'une séquence génomique, et qui se trouve associée à des régions régulatrices, des régions transcrites et d'autres régions fonctionnelles ».

QCM - QROC

1.1 Qu'appelle-t-on génotype ?

1.2 Qu'appelle-t-on phénotype ?

1.3 Quelle est la bonne affirmation ?

a) Un allèle est une forme particulière d'un gène.

b) Un allèle est un gène obligatoirement muté.

c) Un allèle est la forme protéique d'un gène.

1.4 Quelle est la réponse la plus acceptable à la question « Qu'est-ce qu'un gène » ?

a) Une séquence d'ADN codant une protéine.

b) Un concept théorique commode sans réalité physique.

c) Une entité chimiquement variable porteuse d'une information transmissible.

d) Une région définie de la molécule d'ADN, correspondant à une unité héréditaire, localisable au sein d'une séquence génomique, et qui se trouve associée à des régions régulatrices, des régions transcrites et d'autres régions fonctionnelles.



1.5 Des jumeaux bivitellins (issus de deux ovules distincts) sont engendrés par un couple, tous deux hétérozygotes (*Aa*) pour l'allèle *a* récessif de l'albinisme. Quelle est la probabilité pour que les jumeaux possèdent le même phénotype de pigmentation (présence ou absence de pigmentation) ?

1.6 Percevoir l'amertume du phénylthiocarbamide (PTC) est une caractéristique dominante, non liée au sexe, et son contraire, une caractéristique récessive. Une femme qui perçoit l'amertume et dont le père ne la perçoit pas, épouse un homme ayant un phénotype de sensibilité. Ce monsieur a déjà eu, dans une union précédente, une fille insensible au PTC. **a)** Quelle est la probabilité que leur première fille soit insensible ? **b)** Soit sensible ? **c)** Leur premier garçon soit sensible ? **d)** La probabilité que leurs deux premiers enfants, de l'un ou l'autre sexe, soient sensibles ?

1.7 Qu'elle est la probabilité pour que le premier enfant d'un couple sain soit atteint de phénylcétonurie, sachant que l'homme a une sœur atteinte et la femme, son frère ?

1.8 La dystrophie musculaire de Duchenne, une maladie liée à l'X, est une grave myopathie qui touche les garçons seulement. Quelle est la probabilité, pour qu'une femme dont le frère est atteint ait un enfant malade ?

RÉPONSES

1.1 C'est la partie du génome d'un individu qui participe à la détermination d'une caractéristique spécifique, d'un trait phénotypique, de l'individu.

1.2 Le phénotype d'un organisme fait référence à son apparence physique générale et à son organisation mais aussi à l'aspect particulier d'un trait variable, comme la couleur des yeux, le comportement, ou une caractéristique métabolique observable entre les individus d'une espèce.

1.3 **a)**

1.4 **d)**

1.5 La probabilité que se produise l'un ou l'autre de deux événements indépendants (qui s'excluent mutuellement) est la somme de leurs probabilités individuelles. Puisque les génotypes des deux parents est



Aa (*A*, allèle sauvage dominant responsable de la pigmentation et *a*, allèle récessif de l'albinisme, absence de pigmentation), la probabilité du phénotype de pigmentation pour les deux jumeaux bivitellins est de 3/4 × 3/4 = 9/16, et celui de l'absence de pigmentation (albinisme) de 1/4 × 1/4 = 1/16. La probabilité que se produise l'un ou l'autre sera donc de : 9/16 + 1/16 = 10/16 soit 5/8.

1.6 On définit une probabilité comme le rapport du nombre d'arrivées d'un événement sur le nombre d'occasions qu'il a de se produire (nombre d'essais). Si *G* est l'allèle dominant sensible et *g*, l'allèle récessif, insensible, les génotypes respectifs du père et de la mère sont *Gg*. La probabilité de chaque allèle étant de 1/2 et celle du génotype insensible *gg* de 1/4, pour chaque sexe, soit une probabilité de (1/4 × 1/2) = 1/8, pour une première fille insensible. La probabilité du génotype sensible *G*- est de 3/4, soit pour une fille ou un garçon de (3/4 × 1/2) = 3/8. La probabilité que deux événements indépendants se produisent simultanément est le produit de leurs probabilités respectives. La probabilité que les deux premiers enfants, indépendamment de leur sexe, soient sensibles sera donc de 3/4 × 3/4 = 9/16.

1.7 Comme le caractère est porté par le HSA 12 les deux parents peuvent être porteurs sains ou homozygotes dominants. Au vu de leur généalogie, les deux parents doivent être hétérozygotes. La probabilité qu'ils présentent cet état pour chacun est de 2/3 et la probabilité qu'ils présentent cet état simultanément est de 2/3 × 2/3 = 4/9. Un quart de leurs enfants sera atteint de phénylcétonurie. La probabilité combinée qu'ils soient hétérozygotes et que leur premier enfant soit atteint de phénylcétonurie est de 4/9 × 1/4 = 4/36 soit 1/9.

1.8 Puisque le frère est atteint cela signifie que leur mère qui n'est pas malade transmet cependant la maladie et possède le génotype X^AX^a, avec X^a désignant l'allèle défectueux. On suppose que son compagnon est sain et possède donc le génotype X^AY. La probabilité de chaque allèle étant de 1/2, la probabilité du génotype malade X^aY sera de 1/4, et celle qu'un enfant le soit de 1/4 × 1/2 = 1/8.





Mutations et sélection

PLAN

OBJECTIFS

- Étudier l'origine et le rôle des mutations
- Établir les relations entre mutations et sélection
- Éclairer la notion de déséquilibre de liaison
- Distinguer la part respective des gènes et de l'environnement

2.1 ORIGINE DES MUTATIONS

Pour survivre et se reproduire, les organismes vivants doivent répliquer fidèlement leur ADN et le protéger des détériorations physico-chimiques. En effet, des mutations chez l'homme peuvent par exemple dérégler le contrôle strict de la division cellulaire, entraîner la prolifération continue de certaines cellules et former une tumeur cancéreuse. De la même manière si les mutations concernent l'ADN du spermatozoïde ou de l'ovule des parents, des maladies génétiques peuvent en résulter chez l'enfant. Des mécanismes moléculaires de correction et de réparation des défauts de l'ADN minimisent ces risques et agissent à l'instar de filtres, certes imparfaits, pour éliminer les dommages causés à l'ADN par les erreurs de réplication ou les agents mutagènes apportés par les nuisances environnementales.



A

1 2 3 4

Étalement de 10^7 cellules

Incubation pendant 48 h

B

16 1 41 9

Nombre de colonies bactériennes résistantes à l'antibiotique

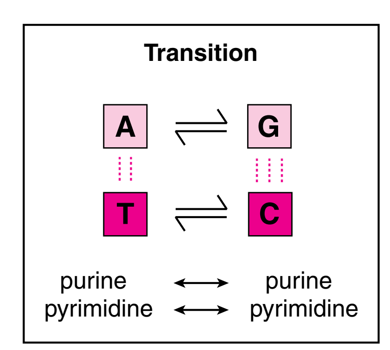
Figure 2.1 Mise en évidence des mutations spontanées.

Les quatre cultures de la même souche bactérienne sont propagées un grand nombre de fois dans un milieu liquide complet puis on réalise à partir de chacune de ces quatre cultures un étalement d'un nombre identique de bactéries sur milieu complet solide en boîte de Pétri **(A)** contenant une même concentration d'un antibiotique puissant. Le nombre aléatoire de bactéries mutantes dans chaque culture **(B)** révèle le caractère spontané et aléatoire des mutations de l'ADN bactérien conférant une résistance à l'antibiotique. Des étalements témoins, non représentés ici, réalisés sur des boîtes de Pétri dépourvues d'antibiotique donnent pour chacune des cultures après 48 h d'incubation un nombre élevé et similaire de colonies. Ceci permet de vérifier que le nombre de bactéries présentes dans les quatre cultures est bien comparable et qu'aucun défaut métabolique dans la croissance des bactéries ne vient perturber l'analyse.

Il faut admettre cependant que si l'ADN était absolument invariant, c'est-à-dire répliqué avec une fidélité absolue et donc dépourvu totalement de mutations, l'évolution des espèces vivantes à l'origine de ce que l'on nomme la biodiversité ne serait pas cette réalité admise par tous aujourd'hui. Les détériorations ou les réarrangements de séquence de l'ADN d'un individu à un instant donné sont donc le résultat de la balance qui s'établit entre les événements de mutations et de réarrangements et l'efficacité des mécanismes de réparation de ces altérations. Les taux de mutations varient selon les espèces. Dans certaines conditions, un taux élevé de mutations peut permettre à certains organismes d'évoluer et donc de s'adapter plus rapidement à leur environnement. Un exemple particulièrement éclairant est fourni par les mutations bactériennes entraînant l'apparition de résistance aux antibiotiques *(Fig. 2.1)*.

Les mutations liées à des erreurs de duplication de l'ADN

Les mutations les plus simples portent sur la modification d'une base seulement. Ce sont les **transitions** et les **transversions** *(Fig. 2.2)*. On les nomme aussi **mutations ponctuelles**. Elles sont à l'origine de la notion de polymorphisme d'un seul nucléotide (SNP pour Single Nucleotide Polymorphism). À ce type de mutations se rattachent aussi les insertions ou les délétions d'un seul nucléotide. Le changement de position d'un atome d'hydrogène dans une base (tautomérisation) peut conduire également à de faux appariements et à des changements de base au deuxième cycle de réplication *(Fig. 2.3)*. Les autres modifications portent sur des fragments d'ADN de taille plus importante. Elles sont engendrées généralement par des événements de **recombinaison** ou de **transposition** qui se rangent parfois dans la catégorie des mutations chromosomiques.



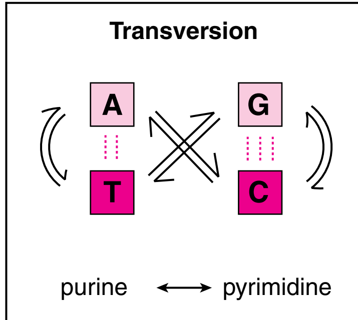


Figure 2.2 Changements de base pyrimidique ou purique par transition ou transversion.

Les taux de mutations varient énormément d'une espèce à une autre. Des taux élevés peuvent permettre dans certaines circonstances à un organisme de faire face à un changement environnemental rapide *(Fig. 2.1)*.

La fréquence des mutations quel que soit leur type varie de 10^{-6} à 10^{-11} par cycle de réplication de l'ADN avec des régions chromosomiques particulièrement instables où la fréquence est très élevée. Elles sont nommées pour cela **points chauds de mutations**.

Il faut ranger dans cette catégorie les séquences répétées de deux à quatre nucléotides, appelées **microsatellites**. Ce sont des répétitions de bases cytosine et adénine sur un brin d'ADN qui s'écrivent $(CA)_n$ ou en sens inverse sur l'autre brin $(TG)_n$. Lors de la réplication de ces microsatellites, des dérapages de la machinerie de duplication se produisent fréquemment, entraînant des variations dans le nombre n de répétitions effectivement copiées. Pour un microsatellite donné, il est donc fréquent d'observer parmi les individus d'une même espèce, un fort polymorphisme du nombre de répétitions (c'est-à-dire de la taille du microsatellite). Compte tenu du nombre très élevé (plusieurs centaines de milliers) de microsatellites dans le génome des eucaryotes, notamment chez l'homme, le choix approprié de certains d'entre eux et la combinatoire qui en découle dans leur polymorphisme de longueur, constitue un moyen technique efficace pour caractériser tout individu de l'espèce concernée. C'est la technique dite des **empreintes génétiques** largement utilisée dans les démarches d'identification des personnes et de leurs caractéristiques génétiques.

Sous sa forme ponctuelle, la mutation nécessite deux cycles successifs de réplication pour se fixer. Elle prend son origine dans l'un des 12 appariements incorrects possibles, lors du premier cycle de réplication de l'ADN. Elle ne devient effective qu'au deuxième cycle de réplication au cours duquel l'appariement incorrect engendre un changement définitif dans l'une des molécules-filles d'ADN *(Fig. 2.3)*.

Les mutations provoquées par des agents mutagènes

Outre les erreurs spontanées de réplication, les mutations sont parfois le résultat de détériorations de l'ADN provoquées par des **agents mutagènes** divers. Les radiations ionisantes et ultraviolettes, certaines substances chimiques (agents alkylants, méthylants, intercalants) sont en cause.

L'hydrolyse spontanée des groupes NH_2 de la cytosine ou de la 5-méthyl-cytosine donne naissance respectivement à l'uracile et à la thymine. Elles s'apparieront alors après un cycle de réplication avec l'adénine et non plus avec la guanine créant ainsi une mutation ponctuelle.

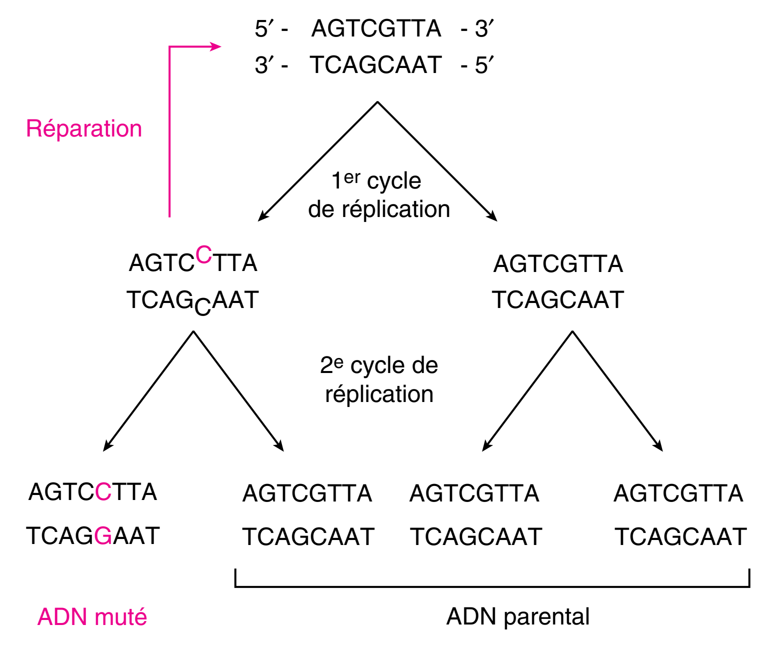


Figure 2.3 Deux cycles de réplication sont nécessaires pour qu'une mutation soit définitivement introduite dans une région de l'ADN.

L'hydrolyse spontanée de la liaison N-glycosidique unissant les bases puriques au désoxyribose crée des pseudo-nucléotides sans guanine ou adénine, dits apuriniques qui bloquent la réplication.

Le transfert par alkylation de groupes éthyl (comme le *N*-éthyl-*N*-nitrosourée ou ENU) ou méthyl (comme l'éthyl-méthanesulfonate ou EMS) sur les bases puriques ou pyrimidiques crée souvent des bases éthylées ou méthylées dont l'appariement naturel est modifié. C'est particulièrement le cas de la O^6-méthylguanine qui peut s'apparier avec la thymine ce qui provoque à la réplication un changement de paires G-C en paires A-T.

Les radicaux libres tels O_2^- ou OH· engendrés à partir de l'oxygène par les radiations ionisantes ou certains agents chimiques attaquent l'ADN et provoquent des **adducts** transformant les bases ciblées en sites hautement mutagènes.

C'est le cas de la 7,8-dihydro-8-oxoguanine (oxoG) qui peut s'apparier aussi bien avec la cytosine que l'adénine, provoquant dans ce dernier cas une des mutations par transversion les plus fréquemment rencontrées à l'origine de certains cancers humains.

La lumière ultraviolette de longueur d'onde voisine de 260 nm (proche UV) est fortement absorbée par les purines et pyrimidines, en

raison de leur système de doubles liaisons conjuguées. Cela peut provoquer tout particulièrement la fusion photochimique de deux thymines adjacentes et donner un **dimère de thymine** *(Fig. 2.4)* qui ne peut pas efficacement s'apparier avec les adénines complémentaires ce qui bloque l'ADN polymérase lors de la réplication.

Les radiations ionisantes (rayons γ et rayons X), seules ou associés à des agents eux-mêmes ionisables en espèces réactives, provoquent des cassures aléatoires de la double hélice d'ADN qui conduisent

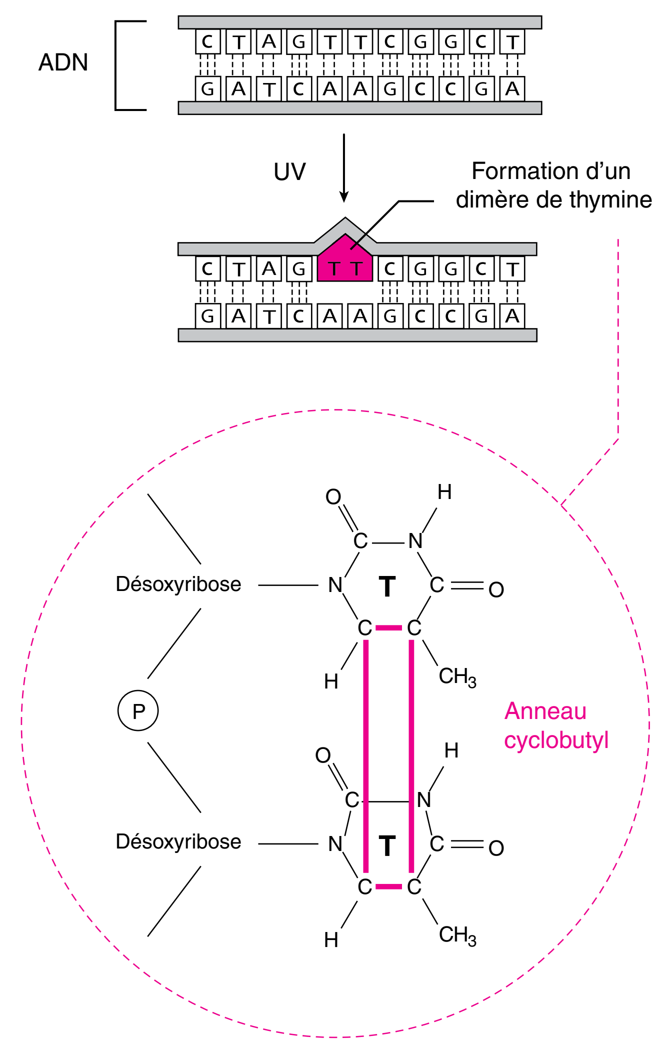


Figure 2.4 Formation et structure d'un dimère de thymine.

souvent à la mort cellulaire. Bien maîtrisé, leur emploi thérapeutique sert à limiter la prolifération des cellules tumorales dans le traitement de certains cancers.

Constitués de molécules polycycliques, les agents intercalants (éthidium, acridine orange, proflavine) ont des formes aplaties mimant les systèmes formés par les purines et pyrimidines hétérocycliques appariées. Ces agents hydrophobes s'intercalent facilement entre les plateaux formés par les bases au sein de la double hélice. En se glissant entre les plateaux, ils les déforment et favorisent, lors de la réplication, des insertions ou des délétions de nucléotides.

Non réparées, toutes ces détériorations interfèrent avec la réplication ou la transcription et mettent en cause la survie cellulaire. Il n'est donc pas étonnant qu'au cours de l'évolution, des systèmes d'identification et de réparation de ces dommages aient été sélectionnés. Ils sont basés sur des mécanismes d'excision de la base ou du nucléotide défectueux, de réparation par recombinaison quand les deux brins de l'ADN sont en cause ou dans les cas extrêmes sur un mécanisme de « passage en force » au travers de la lésion.

La réparation par **excision d'une base** incorrecte utilise une **ADN glycosylase** qui reconnaît et élimine la base indésirable en hydrolysant la liaison glycosidique unissant le pentose à la base. La réparation par **excision d'un nucléotide** est basée sur la reconnaissance de la distorsion de la double hélice occasionnée par le nucléotide incorrect. Lorsque la détérioration concerne les deux brins lors d'une cassure complète de l'ADN, la réparation ne peut pas s'effectuer en utilisant l'un des brins comme matrice modèle. Dans ce cas, un ADN homologue est utilisé comme une matrice modèle de secours, à la condition qu'il soit effectivement présent dans la cellule.

Il arrive que l'ADN polymérase rencontre au cours de la réplication une détérioration non réparée et soit bloquée dans son action de duplication. Un mécanisme de forçage de la lésion peut alors se déclencher. Il est basé sur l'action d'ADN polymérases spéciales, les ADN polymérases de la famille Y, qui synthétisent l'ADN directement en ignorant la lésion, sans le recours à l'appariement conforme à une matrice pour incorporer les nucléotides. La conséquence inévitable de ce type de synthèse est l'introduction fréquente de nucléotides incorrects. Cela assure toutefois la poursuite de la réplication, essentielle pour la survie de la cellule, même au prix d'un taux élevé de mutations.

Contrairement aux idées prédominantes en vogue après la découverte de sa structure, insistant sur le caractère invariant de l'ADN, on considère aujourd'hui la recombinaison et plus largement la mobilité et la plasticité de la séquence, comme des caractéristiques courantes et indispensables au fonctionnement et à la reproduction des êtres vivants.



Coupure simple brin sur chaque molécule

A B C
5′ 3′
3′ 5′
A B C
a b c
5′ 3′
3′ 5′
a b c

Brins homologues alignés

Appariement croisé

A B C
5′ 3′
3′ 5′
A B C
a b c
5′ 3′
3′ 5′
a b c

Jonction de Holliday

Migration de branche

A B C
5′ 3′
3′ 5′
A b
a B
5′ 3′
3′ 5′
a b c

Hétéroduplex

1
2
Clivage
Clivage

A B c
A b c
a B C
a b C

A B C
A b C
a B c
a b c

Duplex commutés et croisés

Duplex commutés

Figure 2.5 Mécanisme de la recombinaison homologue selon le modèle de Holliday. Les lettres ABC et abc désignent des régions homologues des ADN où les séquences sont cependant légèrement différentes. Les duplex hybrides résultant de la résolution de la jonction diffèrent selon le type de clivage.



Cette réalité se manifeste particulièrement à la méïose, notamment lors de l'échange de fragments entiers de chromosomes, par **recombinaison homologue**, dans un processus connu sous le terme de « **crossing over** » (voir chapitre 6). Elle est également illustrée par les réarrangements et les déplacements de séquences connues sous le nom de **transposition**.

La recombinaison homologue existe chez tous les organismes. Son mécanisme fondamental repose sur la rupture suivie de la réunion de deux molécules d'ADN homologues, c'est-à-dire identiques sur des régions d'au moins une centaine de bases, ce qui n'exclut pas la possibilité de différences comme il en existe entre deux allèles d'un même gène. Les deux molécules recombinantes sont porteuses de cassures symétriques sur l'un ou les deux brins et un appariement croisé s'établit entre les brins des deux molécules qui se connectent ainsi en une structure désignée par le nom de son découvreur, la **jonction de Holliday** *(Fig. 2.5)*. Cette structure de jonction migre le long des deux molécules d'ADN par séparation progressive des simples brins. Le processus est appelé **migration de branche**. Il assure une hybridation croisée des brins des deux origines. Finalement, des clivages interviennent. Ils libèrent deux duplex d'ADN ayant échangé des régions de taille variable selon les circonstances. Cette opération se nomme **résolution de la jonction**. Selon le type de clivage *(Fig. 2.5)*, le réassortiment des régions entre les deux ADN concernés par la recombinaison fournit soit des duplex commutés et croisés, soit des duplex simplement commutés.

2.2 LES DIFFÉRENTS TYPES DE MUTATIONS MODIFIANT OU NON LA FONCTION

Les mutations géniques

Ce sont tout d'abord les mutations ponctuelles, comme les insertions, les délétions et les substitutions qui appartiennent à la catégorie des transitions et des transversions (voir *Figure 2.2*). Quand elles interviennent au sein d'une séquence génique codant une protéine, trois principales conséquences peuvent en résulter :

- La mutation sera dite « silencieuse » si l'acide aminé n'est pas modifié ce qui est souvent le cas si le changement de base concerne la 3e position du codon (voir à ce sujet, le code génétique dans le Mini Manuel de *Biologie moléculaire*)
- La mutation sera dite « faux sens » si l'acide aminé est différent de celui normalement codé.
- La mutation sera dite « non sens » si un codon stop est formé, ce qui aboutit à une protéine tronquée.



➤ Les mutations « faux sens » par substitution de bases, ainsi que les mutations « non sens », respectueuses du cadre de lecture, à l'origine de maladies humaines, sont les plus fréquentes (59 %). Les délétions représentent 22 % et les insertions/duplications, 7 %. Les mutations modificatrices des sites d'épissage et celles présentes dans les régions régulatrices des gènes, représentent environ 11 % *(Tableau 2.1)*.

Tableau 2.1 Fréquence relative des types de mutation responsables de maladies humaines.

Mutation	Nombre	% du total
Délétion	6 085	21,8
Insertion/duplication	1 911	6,8
Réarrangement complexe	512	1,8
Répétition	38	0,1
Non sens/faux sens	16 441	58,9
Épissage	2 727	9,8
Régulatrice	213	0,8
Total	**27 027**	**100**

L'effet d'une mutation ponctuelle est parfois annulé par une deuxième mutation qui restaure la base initialement présente (c'est une vraie réversion) ou qui intervient sur un deuxième site grâce auquel la fonctionnalité de la protéine sera restaurée. C'est le cas d'une délétion qui interrompt le cadre de lecture qu'une insertion très proche restaurera. C'est le cas aussi des ARN de transfert suppresseurs de mutations où l'anticodon est porteur d'une mutation qui permet de lire un codon sens préalablement muté en codon stop.

Les insertions peuvent comporter plusieurs bases ou nucléotides. Les éléments transposables (transposons) ou les répétitions de bases (comme AC ou TG) présentes dans les microsatellites et incorrectement répliquées sont de bons exemples de ce type d'événement. Lorsqu'ils se produisent à l'intérieur d'une séquence génique les effets portent sur la séquence codante ou bien les sites d'épissage ce qui dans les deux cas, altère finalement la protéine. Le transposon peut cependant à son tour être excisé, ce qui restaure la séquence initiale. Dans le cas d'une délétion, la séquence codante d'un gène sera irréversiblement altérée *(Fig. 2.6)*.



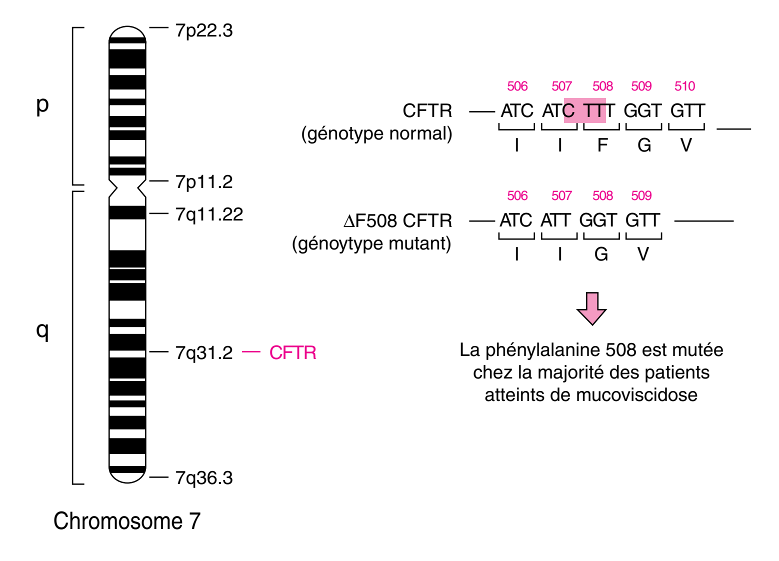


Figure 2.6 Délétion de trois bases dans la séquence ADN du chromosome 7 humain.

Cette délétion très fréquente chez les patients atteints de mucoviscidose, élimine une phénylalanine de la séquence polypeptidique avec comme résultat d'inactiver la protéine qui contrôle le passage des ions chlore à travers la membrane des tissus sécréteurs ce qui bloque leur fonctionnement normal par un encombrement dû aux mucosités et autres sécrétions qui ne sont pas évacuées.

Les mutations chromosomiques : amplifications, délétions, translocations, inversions, perte d'hétérozygotie

Les mutations peuvent intéresser des régions entières de chromosomes. Ce sont des événements qui se produisent au moment de la division cellulaire lorsque des recombinaisons ont lieu, particulièrement au moment de la méiose donnant naissance aux gamètes. Ces derniers transmettront à la descendance la mutation chromosomique dont ils sont porteurs. Comme pour les mutations ponctuelles, les mutations chromosomiques se présentent sous la forme de duplications (ou amplifications) et de délétions de régions entières qui amènent souvent la fusion de gènes à l'origine séparés. Dans ce dernier cas, il peut s'agir de translocations ou d'inversions de fragments entre deux chromosomes. L'événement se trouve souvent à l'origine de fusions géniques dont les produits sont des protéines de fusion, parfois responsables de processus de tumorisation et de cancers *(Fig. 2.7)*. C'est le cas par exemple de la délétion chromosomique conduisant à la fusion des gènes *FIG* et *ROS* et à la production d'une protéine de fusion possédant une

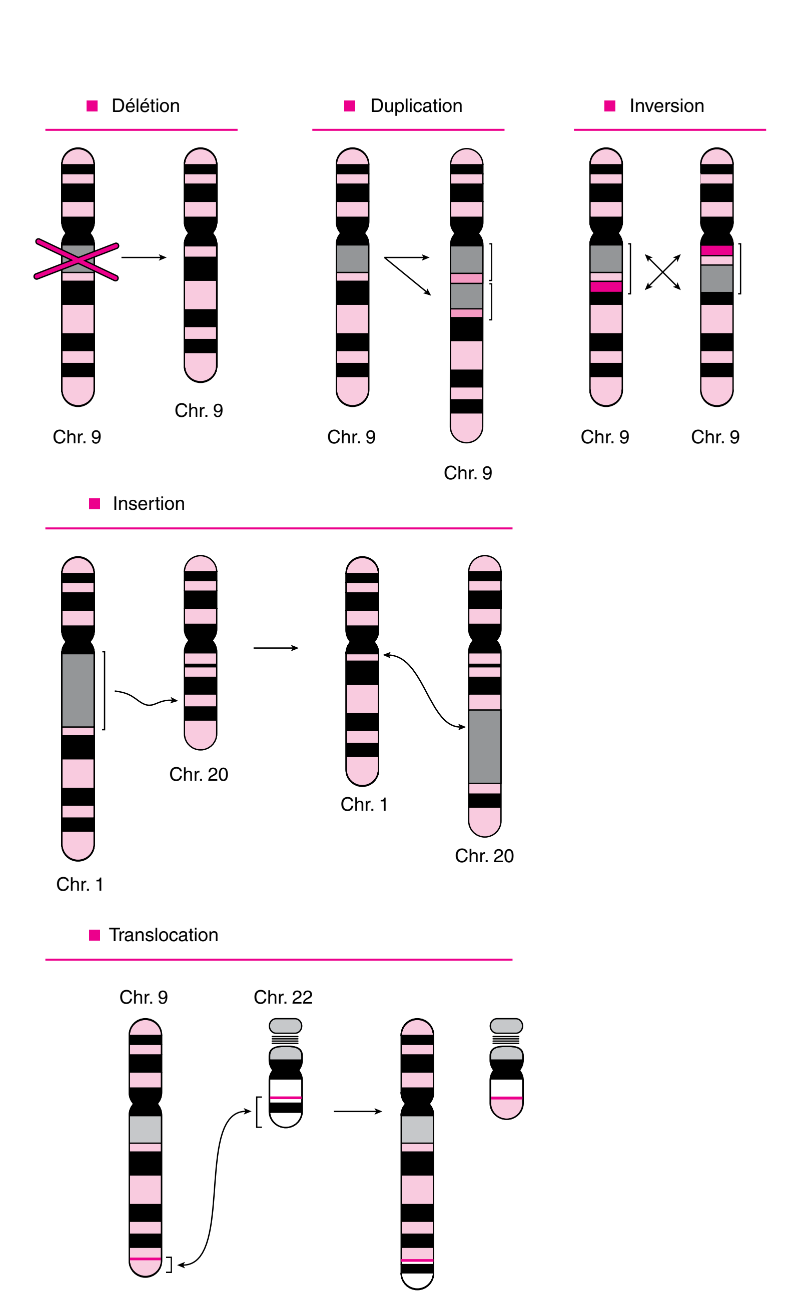


Figure 2.7 Les cinq principaux types de mutations chromosomiques.

activité permanente de tyrosine kinase, ce qui a pour effet la transformation tumorale des cellules gliales et de provoquer un cancer du cerveau.

Dans un organisme qui présente deux allèles différents pour un même gène (état hétérozygote), la délétion ou un événement de recombinaison spécifique peuvent conduire à la perte de l'un des deux allèles et à la perte de l'hétérozygotie.

Les mutations pertes ou gains de fonction

Les effets phénotypiques consécutifs à une mutation sont parfois utilisés pour qualifier la mutation elle-même. C'est ainsi qu'on parle de mutations « **perte de fonction** » ou « **gain de fonction** ». Lorsque la perte de fonction est consécutive à l'inactivation ou la disparition de l'un des deux allèles d'un gène, on parle d'allèle nul et la mutation est dite amorphique (sans phénotype). Très souvent de telles mutations sont récessives, sauf lorsque qu'elles concernent des organismes haploïdes (présence d'un seul génome) ou que le phénotype, pour être exprimé, nécessite l'activité des deux allèles du gène d'un organisme diploïde. Dans ce cas, on parle d'haploinsuffisance pour signaler une expression incomplète du phénotype, due à la perte d'activité de l'un des deux allèles.

Les mutations dites « gain de fonction » modifient le produit du gène. L'ARN ou la protéine acquière une fonction nouvelle qui change généralement le phénotype observé chez l'organisme considéré (mutation néomorphique). Ces mutations sont généralement dominantes.

Il existe un type de mutation dite « **dominante négative** » dont l'action se manifeste par une altération de l'effet de l'allèle de type sauvage responsable du phénotype habituel. Le produit altéré (ARN ou protéine) de l'allèle muté, interagit négativement avec le produit de l'allèle sauvage qui ne peut alors exprimer pleinement son effet phénotypique. Un bon exemple en est donné par le syndrome de Marfan, une maladie génétique humaine due à une mutation du chromosome 15 touchant le gène *FBN1* qui code la fibrilline 1, une glycoprotéine essentielle pour la formation des fibres élastiques des tissus constitutifs des ligaments associés à la formation du squelette et à son équilibre, mais aussi constitutifs de l'aorte et de l'œil. En se liant au facteur de croissance TGFβ, la fibrilline contrôle son niveau d'activité. La mutation du gène *FBN1* réduit le niveau de la fibrilline et un excès de TGFβ en résulte qui est à l'origine du syndrome.

Les mutations dites létales conduisent à la mort de l'organisme porteur. Lorsqu'elles se manifestent au stade embryonnaire précoce, l'embryon meurt et les mutations restent souvent de ce fait, ignorées.

Autres types de mutations : phénotypiques, biochimiques et conditionnelles

Parce qu'elles affectent l'aspect de l'organisme porteur certaines mutations sont dites « morphologiques ». Elles se manifestent par des modifications de la taille ou de l'aspect d'un organe. De telles mutations ne sont pas nécessairement associées à des changements de la séquence génomique. Elles résultent souvent d'altérations épigénétiques dont l'origine se trouve dans l'action de facteurs environnementaux. Parfois cependant un mutant morphologique résulte directement d'une mutation dans l'un des composants d'une voie enzymatique. On parle alors de mutation « biochimique » (voir chapitre 5). Dans d'autres circonstances, ce sont les conditions environnementales qui imposent le phénotype correspondant. Ainsi une mutation sensible à la température provoquera la mort cellulaire si la température s'avère trop élevée (conditions restrictives) alors qu'à plus basse température (conditions permissives) la cellule survivra.

Nomenclature

Il existe une nomenclature simple pour désigner les mutations ponctuelles qui sont des substitutions ou des délétions. On écrira par exemple Q213H pour nommer la substitution de la glutamine en position 213 par l'histidine. Q désigne la glutamine dans le code international à une lettre, 213 désigne la position de la glutamine à partir de l'extrémité N de la protéine et la lettre H, l'histidine qui est substituée à la glutamine.

Dans le cas d'une délétion, on écrit ΔQ213. Le signe grec Δ (delta) indique la délétion. La lettre Q désigne la glutamine présente dans la séquence de type sauvage et le nombre désigne la position où se trouve la glutamine à partir de l'extrémité N de la protéine.

On procède de façon similaire pour désigner les mutations affectant les bases puriques et pyrimidiques de l'ADN et de l'ARN.

2.3 LA NOTION DE PARAMUTATION

Cette notion fait référence à l'existence d'une interaction entre les deux allèles d'un locus. En effet, dans la **paramutation**, un des deux allèles présents à une génération affecte de façon héritable l'autre allèle dans les générations ultérieures, même si l'allèle responsable du changement n'est pas lui-même transmis. La paramutation constitue donc une exception à la règle mendélienne de la ségrégation indépendante des allèles. Quel est le support physique qui est responsable de la transmission de l'information génétique dans un tel phénomène ? La réponse désigne aujourd'hui de petits ARN, sous toutes les formes



mâle *Kit* $^{tm1Alf\,/\,+}$ × femelle *Kit* $^{+\,/\,+}$

ARNm de Kit aberrant

Spermatozoïdes

Ovule

Embryon

F1

mâle *Kit* $^{+\,/\,+}$ × femelle *Kit* $^{+\,/\,+}$

F2

Kit $^{+\,/\,+}$

Figure 2.8 Paramutation du gène *Kit* chez la souris.

À la génération F1, bien que les souris mâles aient un génotype de type sauvage +/+, le phénotype de leur queue correspond à un génotype muté *Kit*^{tm1Alf}/+. Croisés avec des femelles de type sauvage +/+, les individus F2 continuent de présenter, quoique souvent à un moindre degré, le phénotype mutant. L'extrémité de la queue de souris mâles possédant un allèle paramutagénique *Kit*^{tm1Alf} et un allèle de type sauvage (+) est blanche au lieu d'être colorée. L'allèle *Kit*^{tm1Alf} produit un ARNm incorrect qui est embarqué dans le sperme et transmis à l'embryon après fécondation de l'œuf. Même si les descendants possèdent les deux allèles sauvages, ils expriment cependant un phénotype dit « extrémité de la queue blanche ». Ce phénotype est dû à l'action de l'ARNm *Kit*^{tm1Alf} qui provoque soit une dégradation ciblée interférentielle de l'ARNm sauvage, soit un blocage de sa traduction. La disparition progressive au cours des générations successives de l'ARNm *Kit*^{tm1Alf} incorrect ou de ses produits inhibiteurs dérivés conduit progressivement à une perte de la paramutation.

non codantes connues. Ces ARN sont transmis par les cellules sexuelles (gamètes) de génération en génération. L'ARN doit donc être considéré à part entière comme une molécule de l'hérédité.

Des exemples sont connus chez les végétaux. L'allèle paramutagénique R^{st} au locus R peut influencer l'allèle dominant paramutable R^{R}. R^{st} peut même, au long des générations, contrôler un large éventail de phénotypes de coloration chez le maïs, depuis l'absence de couleur des grains jusqu'à la plus intense des colorations. Un tel phénomène est également rencontré chez les mammifères *(Fig. 2.8)*.

2.4 EFFETS DES MUTATIONS

Effets défavorables

La plupart des mutations, notamment celles qui sont ponctuelles ou celles qui ne concernent qu'une région très limitée du génome n'ont pas d'effet phénotypique. Elles sont dites neutres. Les effets indésirables se manifestent par contre dès qu'une mutation de l'ADN, non corrigée par le système de réparation, modifie :

- la séquence d'une protéine au point d'altérer sa fonction moléculaire,
- la quantité du produit du gène ou de la région mutée, qu'elle code une protéine, même si la fonction de celle-ci n'est pas altérée, ou un ARN qu'il soit lui-même codant ou non codant
- la localisation subcellulaire du produit du gène (ARN ou protéine).

Le résultat aboutit à une altération du fonctionnement de la cellule cible de la mutation et ce désordre génétique, générateur d'un effet phénotypique souvent défavorable se transmettra à la descendance si la cellule en question appartient à la lignée germinale (lignée formatrice des gamètes) de l'individu.

Effets favorables

Il arrive cependant que certaines mutations aient des effets positifs et confèrent aux organismes qui en sont porteurs des propriétés qui améliorent leur adaptation aux changements environnementaux. Deux exemples concernant l'espèce humaine sont particulièrement démonstratifs de ce phénomène. L'un des cas concerne les personnes porteuses d'une mutation ponctuelle de la globine (un acide glutamique est substitué par la valine) responsable de l'anémie falciforme (l'agrégation de l'hémoglobine déforme les hématies qui adoptent une forme en faucille et circulent mal) mais qui sont par contre en raison de cette mutation résistante à la malaria (paludisme), l'agent infectieux (*Plasmodium*) étant mal transporté.

L'allèle responsable est un allèle autosomal (chromosome 11) récessif. À l'état hétérozygote, les porteurs de la mutation ne présentent que



quelques hématies déformées et ne sont pas anémiés, mais cela suffit pour leur conférer la résistance à la malaria. C'est un cas où l'hétérozygotie d'un allèle confère au porteur un avantage sur les deux génotypes homozygotes.

L'autre cas concerne une délétion de 32 bases dans le gène codant un récepteur de cytokines nommé CCR5 qui est exprimé préférentiellement dans les lymphocytes T, les macrophages et les cellules dendritiques. Ce récepteur est utilisé par le VIH (virus du sida) comme corécepteur pour coloniser les lymphocytes T. Sous sa forme mutée (CCR5-Δ32) il est non fonctionnel et confère aux porteurs homozygotes une résistance à l'infection par le VIH ; pour les hétérozygotes, il retarde le processus infectieux. Cette mutation, plutôt rare, n'affecte pas la santé des porteurs qui sont plus fréquents dans les populations européennes où sa présence serait liée à une sélection naturelle intervenue vers le milieu du XIVe siècle vis-à-vis de la peste bubonique qui sévissait alors en Europe. En effet, l'agent bactérien de la peste (*Yersinia pestis*) utilise également CCR5 pour coloniser les cellules du système immunitaire.

2.5 LA NOTION DE SÉLECTION

La sélection est un processus biologique complexe à la base de l'évolution des espèces. Il se déroule dans la nature (sélection naturelle) mais peut aussi être dirigé à son profit par l'homme (sélection artificielle).

La sélection naturelle

La sélection, sous sa forme naturelle théorisée en 1859 par Charles Darwin, aboutit à ce que les caractéristiques (traits phénotypiques) favorables, transmises entre individus capables de se reproduire deviendront de plus en plus fréquentes chez les descendants au cours des générations successives. Si les caractéristiques ainsi sélectionnées ont une base génétique, au sens des découvertes faites à la même époque par Gregor Mendel, le génotype correspondant deviendra également de plus en plus fréquent dans les générations suivantes. De façon symétriquement inverse, la fréquence des caractéristiques défavorables et celle de leurs génotypes correspondants diminueront progressivement.

La sélection naturelle peut se manifester à divers stades du cycle vital d'un organisme ayant une reproduction sexuée *(Fig. 2.9)*.

Il arrive que la sélection naturelle de leurs caractéristiques phénotypiques assure aux individus porteurs une excellente adaptation au milieu dans lequel ils vivent (niche écologique). Le maintien d'une telle sélection basée sur un processus interactif entre génotype et milieu peut aboutir finalement à l'émergence d'une nouvelle espèce.



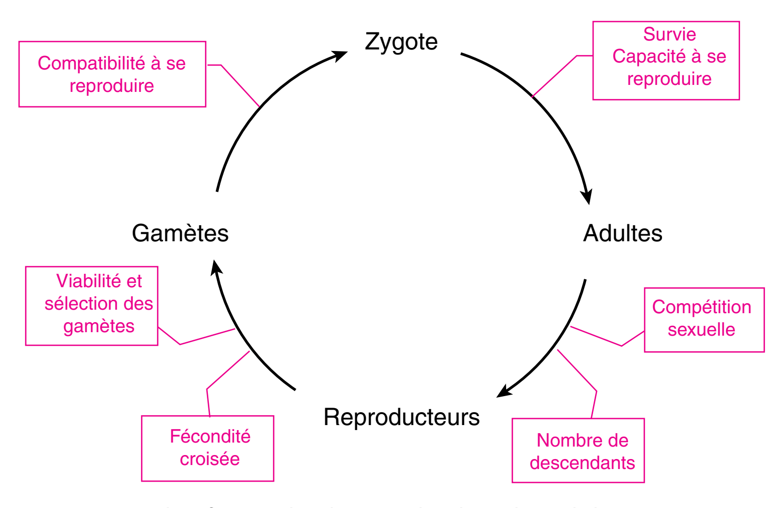


Figure 2.9 Identification des divers stades du cycle vital d'un organisme où s'exerce avec des effets majeurs la sélection naturelle.

Capacité à se reproduire (ne pas disparaître avant d'être capable de se reproduire), **Compétition sexuelle** (compétition entre adultes pour se reproduire), **Nombre de descendants** (et durée de la période d'aptitude à la reproduction), **Fécondité croisée** (fécondité relative mâle et femelle), **Viabilité et sélection des gamètes** (suivant les événements de recombinaison méïotique), **Compatibilité des gamètes** (plus ou moins grande compatibilité sperme-ovule).

Cette idée simple repose sur le concept d'**adaptation** et demeure aujourd'hui encore à la base de la théorie de l'évolution des espèces vivantes, même si celle-ci s'enrichit depuis les années 1950, de nombreuses connaissances cellulaires et moléculaires.

On retrouve le concept d'adaptation (et d'**adaptabilité**) dans la célèbre expression décrivant la sélection naturelle comme « la survie du plus apte », expression que l'on remplace volontiers aujourd'hui par « le plus apte est celui qui peut le mieux se reproduire ». En effet, si un individu vit deux fois moins longtemps que les autres de la même espèce, mais qu'il engendre deux fois plus de descendants capables eux-mêmes de se reproduire, ses gènes seront bien mieux représentés dans les générations suivantes que ceux de ses congénères.

Le concept d'adaptation d'un organisme dépend donc en première approximation de l'adaptabilité des allèles (gènes) qui l'ont façonné. Mais, l'adaptabilité d'un individu particulier se réfère habituellement à un trait ou un ensemble de traits phénotypiques. Il est alors préférable de relier plus simplement l'adaptabilité de l'individu à l'adaptabilité de son génotype. On ne doit pas oublier cependant que l'adaptabilité exprime la valeur sélective moyenne d'une population et non la valeur sélective d'un individu de cette population. C'est ainsi, en appui de



cette observation, qu'un allèle porteur d'une mutation favorable, apparue chez un individu qui cependant disparaît accidentellement avant de pouvoir se reproduire, aura une adaptabilité sélective nulle, bien que le trait phénotypique résultant ait pu être bénéfique pour l'organisme en question.

Finalement, la sélection naturelle peut aboutir à ce que la **fréquence** d'un allèle augmente parce que son adaptabilité sera plus grande que celle des autres allèles. Il peut arriver même que l'allèle soit fixé, c'est-à-dire que sa fréquence soit de 1 (aucun autre allèle n'est présent dans la population). Il peut arriver à l'inverse que la sélection maintienne à un faible niveau un allèle ayant un effet défavorable sur le phénotype, le processus pouvant aller parfois jusqu'à l'élimination de ce dernier (fréquence 0). La situation la plus courante cependant est celle du maintien d'un allèle avec une fréquence intermédiaire dans la population. C'est souvent le cas des organismes diploïdes où les individus hétérozygotes porteurs de deux allèles différents (un sur chaque chromosome homologue) possèdent une meilleure adaptabilité que les individus homozygotes porteurs de deux allèles identiques. L'un des exemples les mieux connus de ce processus est celui des porteurs de l'allèle du gène de la globine conférant une résistance à la malaria. Cet allèle, à l'état homozygote, est par ailleurs responsable de l'anémie falciforme. Dans cette maladie, l'hémoglobine s'agrège de telle manière que les globules rouges se déforment en forme de faucille ou de croissant et circulent très mal dans les vaisseaux, d'où une anémie qui peut être fatale. Les individus hétérozygotes possèdent cependant suffisamment de globine normale et ne sont pas anémiés ; ils sont, par ailleurs, résistants à la malaria. De façon étonnante, l'hétérozygotie leur confère une survie et une capacité à se reproduire supérieure aux deux formes d'homozygotie. D'autres exemples de sélection de l'hétérozygotie sont connus, les plus inattendus étant ceux qui semblent responsables ou associés à la sélection avantageant la parentèle et les comportements d'altruisme réciproque.

Dans sa version la plus étonnante, l'**altruisme** consiste pour un organisme (unicellulaire ou multicellulaire) à se « sacrifier » afin d'apporter à ceux qui lui sont les plus proches génétiquement, sa fratrie en quelque sorte, un secours essentiel à leur survie, avec comme conséquence, à défaut de promouvoir la diffusion de ses propres gènes, de favoriser la diffusion de gènes apparentés. De nombreux exemples ont été identifiés et interprétés avec cette grille de lecture. Ils vont des bactéries et protistes jusqu'aux mammifères et à l'homme. Lorsque les conditions environnementales deviennent défavorables (manque de nourriture) les myxobactéries et les myxomycètes qui vivent normalement à l'état unicellulaire, s'agrègent, forment des colonies où certaines cellules

meurent afin d'assurer la survie des autres. Ce type de comportement assimilable à un comportement social se retrouve chez le nématode *C. elegans* où une mutation ponctuelle dans un neuropeptide détermine le comportement alimentaire, individuel ou de groupe. Il est rencontré aussi chez les mammifères où certains individus n'hésitent pas à se désigner aux prédateurs par un signal d'alerte, favorisant ainsi la fuite des autres individus qui lui sont apparentés. Certains auteurs tirent de ces observations des arguments en faveur de l'idée que les mécanismes assurant la survie et la propagation des gènes sont particulièrement efficaces du point de vue de l'évolution. Certains autres auteurs vont jusqu'à considérer que les comportements solidaires et charitables de certains individus de l'espèce humaine ne repose pas sur des facultés morales supérieures, fruits du progrès de la civilisation et correctrices d'une sorte égoïsme génétique, mais plutôt sur des circuits neuronaux primitifs, habituellement activés dans les relations de satisfaction et de plaisir procurées par la nourriture, l'activité sexuelle ou la relation sociale.

La sélection artificielle

C'est la sélection pratiquée par les cultivateurs, les éleveurs et les sociétés ou organismes producteurs de semences végétales et animales. Elle repose sur le fait que la plupart des traits phénotypiques des espèces vivantes exploitées par l'homme dépendent, lorsqu'ils ont une base héréditaire, de l'interaction de nombreux gènes. Les variations de ces traits, lorsqu'elles sont dues à un seul gène, ne produisent donc à chaque étape de sélection qu'un effet quantitatif très faible sur les phénotypes. Dans une espèce, une race animale ou une variété végétale, les phénotypes d'intérêt se présentent ainsi sous la forme de séries continues au sein desquelles les sélectionneurs cherchent par des accouplements dirigés, à stabiliser ou à éliminer, les traits qu'ils jugent favorables ou défavorables. Ils opèrent généralement en procédant à une sélection par troncature *(Fig. 2.10* et *2.11)*. Tous les individus ayant une valeur pour le trait d'intérêt supérieure à une valeur seuil, sont sélectionnés et utilisés comme reproducteurs. La valeur S ou écart sélectif est définie comme l'écart imposé entre la valeur moyenne du caractère pour la sous-population des reproducteurs sélectionnés à la génération d'origine et la valeur déterminée pour l'ensemble de la population parentale (génération 1). La réponse R à la sélection ou progrès génétique représente la différence entre les moyennes établies pour deux générations successives. Elle montre que la supériorité génétique des individus sélectionnés a été transmise à leurs descendants. Le rapport R/S (réponse de sélection sur écart sélectif) mesure l'augmentation de la valeur additive, c'est-à-dire l'**héritabilité** du caractère considéré.

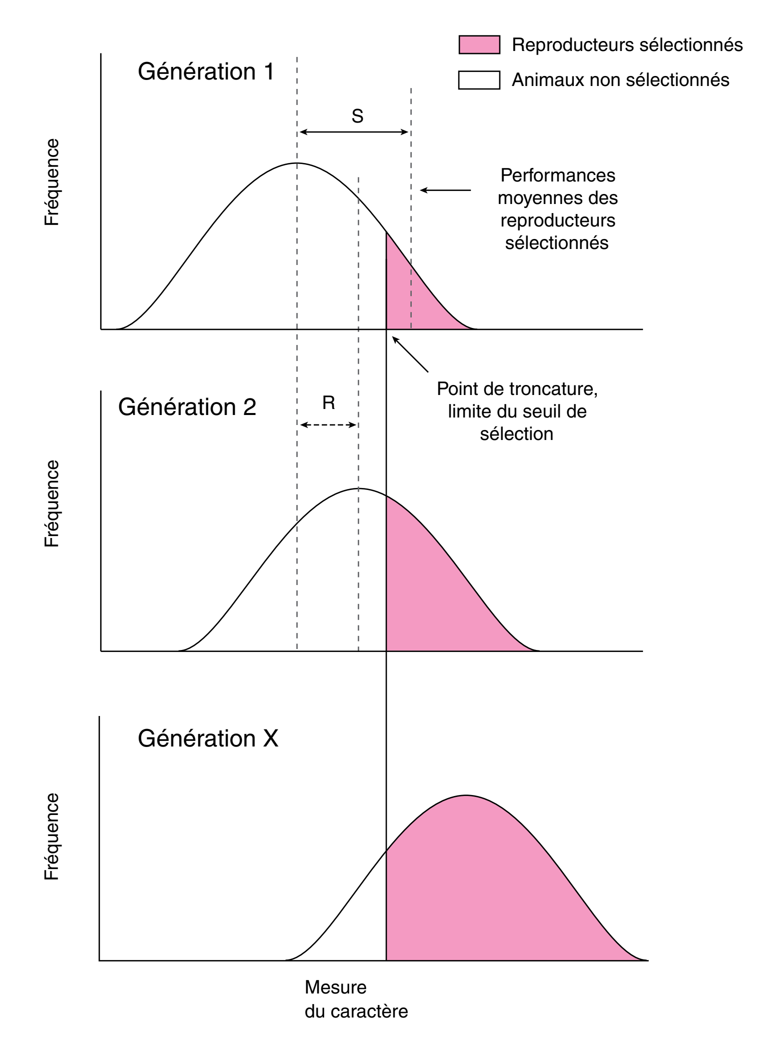


Figure 2.10 Principe de la sélection par troncature après établissement de la courbe de distribution du caractère que le sélectionneur veut améliorer.

La génération 2 résulte d'un croisement entre reproducteurs, sélectionnés à la génération 1. On observe un progrès génétique R, en réponse à la sélection. S mesure l'écart sélectif. En conservant le même point de troncature au cours des générations successives, on enregistre à la génération x, un déplacement très significatif de la courbe de distribution, vers la droite. Cela signifie que le progrès génétique, pour le caractère concerné, est devenu pour la majorité de la population, supérieur au point de troncature, fixé en début de sélection.

Tant qu'une variabilité génétique existe entre les individus d'une population soumise à la sélection artificielle, un progrès génétique peut être enregistré. À partir d'un certain nombre de générations, la **variance génotypique additive** du caractère comme son héritabilité seront nulles. Tous les individus sélectionnés au fil des générations auront théoriquement le même génotype, dû à la fixation des allèles responsables de l'accroissement de la performance pour le caractère choisi.

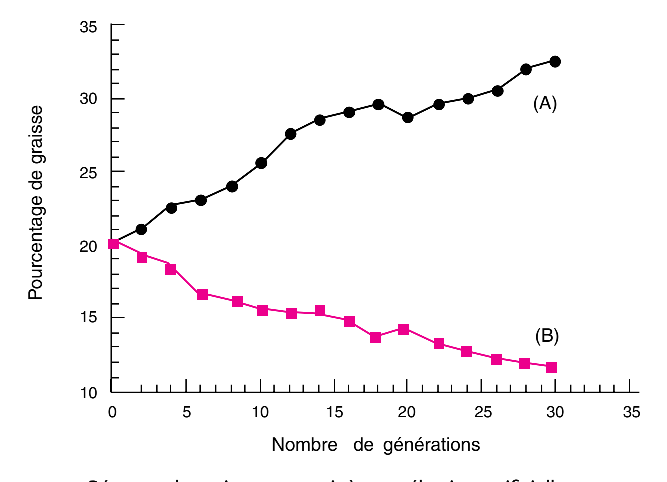


Figure 2.11 Réponse des animaux soumis à une sélection artificielle par troncature.

Les reproducteurs sont sélectionnés par troncature aux deux extrémités de la courbe de distribution du caractère (voir *figure 2.10*). Au fil des générations, cette sélection produit des populations qui divergent fortement pour le caractère concerné (ici le poids de graisse ramené à la carcasse). L'application systématique d'une telle sélection au fil des générations produit des populations A et B qui divergent fortement pour le caractère considéré.

Relations entre sélection et variation génétique

Il doit être noté qu'une partie importante de la variabilité génétique (la différence de génotypes) observée entre les individus s'avère sans effet sur les phénotypes et n'influence pas non plus l'adaptabilité des individus. On dit de cette variabilité associée à la diversité qu'elle est en vérité neutre du point de vue de l'évolution selon la théorie proposée par Motoo Kimura. Autrement dit il n'y a pas de relation directe entre la variation génétique et des différences dans l'adaptabilité des individus constitutifs d'une population. Les mutations qui interviennent dans le génome sans modifier son expression seraient sans effet sur cette adaptabilité. On a pensé pendant longtemps que les mutations neutres concernaient principalement l'énorme fraction des génomes qui



n'est pas occupée par les séquences codantes des gènes à l'origine des protéines. Que des mutations interviennent dans ces régions n'avaient pas d'importance puisque les protéines à l'origine des phénotypes n'étaient pas concernées. Cette conception fait place maintenant à une vision très sensiblement différente depuis que l'on sait que les régions non codantes de l'ADN sont majoritairement transcrites et que les ARN non codants influencent fortement l'expression des gènes et donc les phénotypes. Il est à prévoir que de nouvelles théories à propos de la sélection naturelle et de ses rapports avec l'évolution, intégrant les données concernant la nature des gènes et leur expression verront le jour dans les prochaines années. La conception que l'on se fait de la balance entre mutations et sélection peut s'en trouver modifiée. La sélection est en effet un processus qui en éliminant les mutations défavorables à une adaptation environnementale réduit la variabilité génétique apportée par les mutations. Dans le même temps, de nouvelles mutations, aléatoires, se produisent qui réalimentent la variabilité génétique au sein d'une population. Une balance, un équilibre s'établit entre les deux processus (vitesse d'apparition des mutations et élimination des mutations défavorables). De cette balance résulte en quelque sorte la force propulsive de l'évolution.

2.6 LA DISTINCTION ENTRE PHÉNOTYPE ET GÉNOTYPE

Le phénotype d'un être vivant correspond à son apparence physique d'ensemble ou bien à un aspect particulier (on dit aussi une caractéristique, un trait) mesurable de cette apparence générale, comme sa taille, son poids, sa couleur, une forme, une propriété, une aptitude ou un comportement qui varient suivant les individus examinés. Pour une large part, le phénotype dépend du génotype, autrement dit des gènes et surtout de leurs allèles présents dans les chromosomes de l'individu. La plupart des phénotypes sont déterminés par de nombreux gènes et sont influencés par les facteurs physico-chimiques et biologiques de l'environnement dans lequel vit l'individu ou la population à laquelle il appartient.

La description du phénotype d'un individu s'avère en général plus facile à réaliser – par de simples mesures structurales, biochimiques, physiologiques ou comportementales – que celle de son génotype qui nécessite d'accéder à la séquence de l'ADN. On ne sait en effet établir cette séquence que depuis quelques dizaines d'années seulement. Auparavant les généticiens déduisaient la fonction des gènes à partir du seul examen des phénotypes. La génétique classique ou génétique directe part donc des phénotypes pour aller vers les génotypes, à l'aide souvent de mesures et d'analyses quantitatives portant sur les phénotypes (génétique quantitative). Avec les progrès de la biologie moléculaire,

une génétique inverse se développe qui part des génotypes et parfois des génomes entiers, pour connaître les phénotypes. Adeptes de la première génétique, cultivateurs et éleveurs, intègrent aujourd'hui les démarches de la génétique moléculaire dont la puissance prédictive des phénotypes s'accroît avec les progrès réalisés dans la connaissance approfondie des génomes entiers (séquençage et génotypage).

Les différences constatées entre les phénotypes individuels sont révélatrices de la variabilité génétique existant entre les individus. On l'a vu, la sélection naturelle agit sur les phénotypes en favorisant ou non les individus dont les phénotypes assurent la plus ou moins bonne adaptabilité à l'environnement. Sans différences phénotypiques, il ne pourrait y avoir de différences dans l'adaptabilité et par voie de conséquence pas de sélection naturelle et pas d'évolution.

Le jeu complexe qui s'établit entre phénotype, génotype et milieu est parfois représenté par la relation :

génotype + environnement → phénotype

ou, mieux encore afin de tenir compte des variations aléatoires, par la relation :

génotype + environnement + variation aléatoire → phénotype

Ces relations expriment d'une manière simpliste les interactions qui s'établissent entre les gènes et l'environnement. Elles expriment également les termes du débat classique et toujours d'actualité entre ce qui appartient à l'inné (héritable par les gènes) et ce qui appartient à l'acquis (apporté ou non par l'environnement). Dans l'exemple de la phénylcétonurie (voir chapitre 1) un régime alimentaire approprié, dépourvu de phénylalanine suffit à éviter que la maladie génétique se manifeste. Il s'agit là d'un exemple classique des relations entre gène (muté) et environnement (alimentation). Il existe aussi des circonstances où des organismes porteurs d'un génotype particulier agissent de telle manière qu'ils modifient ou choisissent leur propre environnement. Dans ce cas, il s'agit d'un processus inverse de celui, illustré par l'exemple de la phénylcétonurie. L'environnement est créé par le génotype. De très nombreux exemples sont donnés par les comportements de diverses espèces comme, par exemple ceux des populations humaines dans la mise en œuvre de leurs habitats.

2.7 LIAISON GÉNÉTIQUE ET DÉSÉQUILIBRE DE LIAISON

Liaison génétique et sélection

Contrairement aux observations de G. Mendel, certains gènes ne ségrègent pas à la méiose de façon indépendante. Quand les locus de deux gènes sont liés ou très proches l'un de l'autre sur le même chromosome, on dit qu'existe une liaison génétique. Au cours de la formation

des gamètes, un brassage du matériel génétique s'opère grâce à des croisements par recombinaison (« crossing over »). Plus les gènes sont proches les uns des autres sur un chromosome, plus faible est la probabilité qu'un événement de recombiaison entre chromosomes vienne à les séparer. Il en résulte que si l'un des gènes est l'objet d'une sélection (naturelle ou artificielle), les autres gènes qui lui sont proches ont toutes les chances d'être eux aussi sélectionnés. Au final, la sélection agit sur le potentiel de variation du génome et suivant les circonstances et l'échelle de temps considérée, elle peut ou non appauvrir la variabilité intrinsèque du génome, objet de la sélection.

Par une sélection systématiquement dirigée, il peut arriver qu'un gène (un allèle de ce gène), responsable d'un phénotype recherché par l'éleveur ou le cultivateur, voit augmenter sa fréquence au sein de la population. Les allèles qui lui sont liés au locus chromosomique qu'ils occupent verront du même coup également leur fréquence augmenter, quelle que soit leur action effective (neutre, positive ou négative). Poussée à l'extrême, une telle sélection conduira à fixer dans la population et pour la région génomique considérée, un seul haplotype, c'est-à-dire une seule combinaison allélique (l'allèle du gène cible de la sélection et les allèles des gènes liés voisins).

Liaison génétique entre deux caractères présents dans un famille humaine

Il s'agit ici de la liaison génétique entre deux caractères phénotypiques exprimés ou portés par le chromosome X humain *(Fig. 2.12)*. Le premier caractère est un retard mental (représenté sur la figure par un

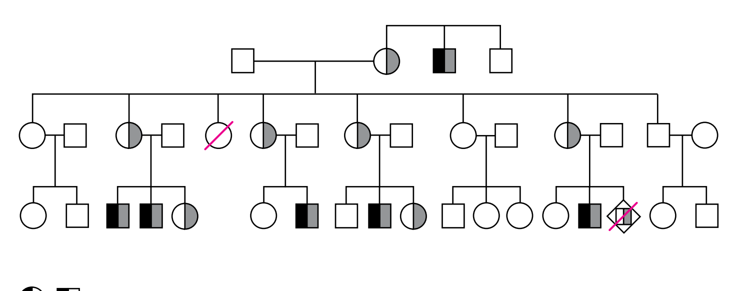


Figure 2.12 Généalogie d'une famille humaine où s'observe chez les garçons une liaison génétique entre un retard mental et la présence d'un site fragile sur le chromosome X.

Le symbole en losange barré correspond à la détection du site fragile lors d'un diagnostic prénatal.

symbole mâle ou femelle à moitié noir, voir *Fig. 1.7* pour la signification des symboles). Le second caractère se rapporte à un aspect particulier du chromosome X qui peut présenter un site dit « fragile », une amorce de cassure souvent visible, chez les individus porteurs, dans les cellules en métaphase (représenté ici par le grisé de la partie droite des symboles).

Le retard mental est récessif et n'est donc pas détecté chez les mères « transmetteuses ». En effet bien qu'elles soient porteuses d'un chromosome X avec un site fragile, et que celui-ci soit détecté sans ambiguïté, elles ne présentent pas de retard mental puisque leur deuxième chromosome X est normal. En examinant les garçons de la famille, on constate que les deux caractères (retard mental et site fragile du chromosome X) sont totalement corrélés. Cela signifie que les déterminants génétiques de ces deux caractères sont très proches sur le même chromosome (en l'occurrence ici le chromosome X). Le chromosome Y des garçons, d'une taille réduite et avec peu de gènes (voir chapitre 3), ne comporte pas les déterminants génétiques (les allèles) du chromosome X. Il ne compense donc pas l'absence du deuxième chromosome X.

Le déséquilibre de liaison mesure une distribution non aléatoire de marqueurs génétiques

Le **déséquilibre de liaison** décrit la situation qui dans une population résulte de certaines combinaisons d'allèles ou de marqueurs génétiques dont la fréquence diffère de celle qu'on attend d'une formation aléatoire d'haplotypes. Autrement dit, les associations non aléatoires entre gènes (allèles) à différents loci sont mesurées par le degré de déséquilibre de liaison.

L'étude des liaisons entre gènes conduit souvent à désigner une région chromosomique de quelques mégabases comme porteuse d'un locus associé à un trait phénotypique d'intérêt (désigné parfois par le terme **QTL**, pour *Quantitative Trait Loci*). Il s'agit souvent chez l'homme du locus où se trouvent les déterminants d'une maladie génétique ; chez l'animal ou la plante, d'un locus associé à un caractère phénotypique intéressant le sélectionneur. Puisque plusieurs centaines de gènes sont susceptibles d'être localisés dans cette région chromosomique, analyser chacun d'eux s'apparente à une mission impossible. L'alternative peut consister non pas à établir l'existence d'une liaison génétique entre un marqueur polymorphe et la maladie ou le QTL, par l'étude de quelques générations d'individus au sein d'un petit nombre de familles, mais au contraire à montrer la coexistence d'un polymorphisme de marqueurs avec le QTL ou la maladie. On réalise pour cela des études d'association au sein d'une population entre des marqueurs polymorphes et le gène fonctionnel (ou le déterminant génétique)

responsable du caractère phénotypique analysé (maladie ou trait quantitatif intéressant). Le raisonnement est basé sur le fait qu'en raison de la grande proximité physique entre le marqueur et le gène (ou le déterminant génétique) un grand nombre de générations sont nécessaires pour qu'ils soient séparés par recombinaison. On s'attend donc à ce que les individus porteurs du caractère analysé présentent le même haplotype de marqueurs avec une fréquence plus élevée que celle qui est attendue d'une distribution aléatoire, même si les individus sont apparentés. Aujourd'hui les marqueurs polymorphes les plus informatifs

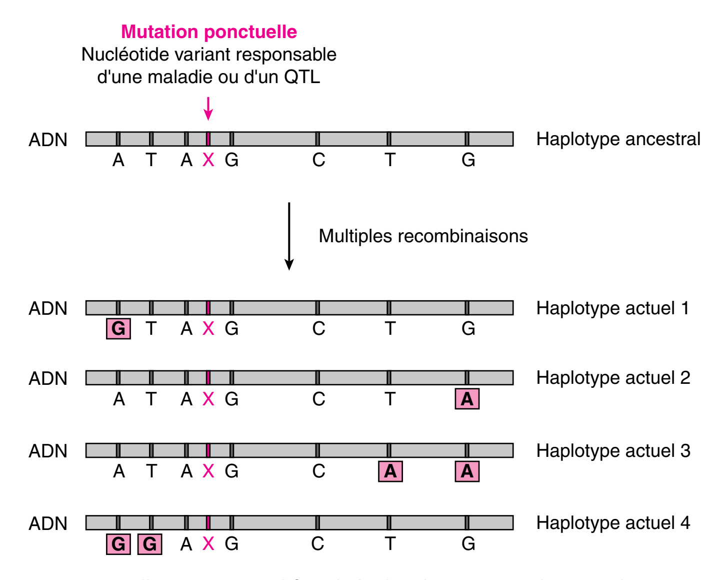


Figure 2.13 Illustration simplifiée de la distribution non aléatoire de marqueurs SNP dans une région chromosomique où doit se trouver une mutation responsable d'un trait phénotypique caractéristique de tous les individus analysés.

On suppose dans l'exemple que cette mutation est responsable d'un trait quantitatif (QTL) chez l'animal ou la plante ou d'une maladie chez l'homme et qu'elle s'est produite antérieurement dans une région caractérisée par l'haplotype ATAGCTG. Elle est désignée ici par la lettre **X**. Au cours des générations successives cet haplotype a subi des recombinaisons. Ainsi, à titre d'exemple, la formation de l'haplotype 1 actuel résulte d'un événement de recombinaison qui s'est produit entre les 1er et 2nd SNP. Les nouveaux allèles qui se sont ainsi progressivement formés au fil des générations sont identifiés dans un carré de couleur rose. La plus petite signature haplotypique conservée (AGC) entre tous les individus porteurs du phénotype d'intérêt, c'est-à-dire en déséquilibre de liaison, montre que la mutation responsable du QTL ou de la maladie pourrait bien se situer entre les SNP 3 et 4 qui sont les plus proches l'un de l'autre et les moins susceptibles d'être recombinants. Cette démarche de cartographie des haplotypes SNP permet ainsi d'identifier une région candidate de taille limitée contenant la mutation d'intérêt dont l'identification définitive sera rendue accessible par exemple par des reséquençages en série.

dans la recherche des régions génomiques associées à des traits phénotypiques d'intérêt sont les **SNP** (Single Nucleotide Polymorphism) c'est-à-dire les nucléotides dont la nature de la base purique ou pyrimidique (A, T, G, C) varie (est polymorphe) selon les individus examinés *(Fig. 2.13)*. La coexistence chez un même individu de deux génomes parentaux conduit au fait qu'un SNP peut se présenter sous un état hétérozygote (les deux bases des deux génomes sont différentes) ou sous l'un des deux états homozygotes (les deux bases sont identiques, chaque paire correspondant à l'une ou l'autre des bases de l'état hétérozygote correspondant).

L'exemple de la *figure 2.13* montre que le déséquilibre de liaison est dû à plusieurs facteurs parmi lesquels on peut citer :

- la distance génétique entre loci ;
- la distance chronologique, c'est-à-dire l'ancienneté de la mutation qui a généré le polymorphisme.

Le déséquilibre de liaison peut s'exprimer par la formule suivante où D_{obs} représente la valeur mesurée du déséquilibre de liaison, D la valeur maximale théorique du déséquilibre de liaison et r le taux de recombinaison :

$$D_{obs} = (1 - r)D$$

Chez les organismes qui se reproduisent de façon asexuée et où par conséquent n'existe pas en principe de recombinaison génétique entre chromosomes parentaux (r = 0), le déséquilibre de liaison est maximal et égal à 1 (tous les gènes sont liés et transmis ensemble sans recombinaison).

2.8 HÉRÉDITÉ ET HÉRITABILITÉ

La notion d'hérédité désigne le processus de transmission des caractéristiques biologiques d'un parent à sa descendance. Les gènes ayant comme support chimique principal l'ADN, l'hérédité désigne par conséquent le processus de transmission des instructions génétiques portées par cette molécule, d'un parent à sa descendance. Certes, ces instructions (l'inné) ne sont pas les seules causes des caractéristiques biologiques d'un individu puisque l'on sait que des facteurs environnementaux (l'acquis) peuvent aussi tout au long de la vie les moduler. Mais l'hérédité c'est la part génétique au sens strict.

La notion d'héritabilité fait référence précisément à l'estimation de cette part génétique en ce qu'elle peut expliquer les différences phénotypiques constatées entre individus d'une même population. L'héritabilité évalue en quoi la variabilité génétique des individus constituant une population peut expliquer la variabilité phénotypique de ces mêmes individus. Plus la variabilité d'un trait phénotypique dépend d'un large

spectre d'instructions géniques (de gènes) plus son héritabilité au regard de chacune des instructions géniques, sera considérée comme faible. À quoi s'ajoute également la part des facteurs environnementaux dans la variabilité phénotypique elle-même.

La relation entre gènes et environnement

Évaluer ce qui dans la formation d'un organisme relève des parts respectives des gènes et de l'environnement (l'inné et l'acquis, la nature et la culture) demeure une question centrale de la génétique. On sait aujourd'hui, dans ses grandes lignes quelles sont les bases moléculaires de l'inné. Les progrès accomplis dans la connaissance de la structure des génomes et de leur fonctionnement commencent également à dégager quelques pistes pour comprendre comment les facteurs de l'environnement agissent sur les génomes pour moduler leur expression et ajouter les multiples degrés de liberté qui leur sont propres. Une nouvelle dimension de la génétique désignée par le terme d'épigénétique se construit peu à peu.

Des aspects déjà anciens de la génétique se sont cependant efforcés d'appréhender dans leur globalité et dans leur diversité, en s'appuyant essentiellement sur des approches mathématiques et statistiques les relations entre gènes et environnement. Il s'agit de ce qu'on appelle la génétique quantitative et la génétique des populations.

La génétique quantitative

La génétique quantitative étudie les caractéristiques biologiques (les traits phénotypiques) qui sont mesurables et quantifiables en essayant de relier la série des valeurs constatées et leur distribution (la variabilité des phénotypes) aux mécanismes sous-jacents tels qu'ils sont admis par une majorité de généticiens se référant aux bases théoriques actuelles de leur discipline (la génétique mendélienne, voir cependant à ce sujet le concept controversé de gène, chapitre 1).

La valeur phénotypique (P) d'un individu résulte de la combinaison d'une valeur génotypique (G) et d'un effet environnemental (E) :

$$P = G + E.$$

La valeur génotypique réunit toutes les instructions géniques ainsi que leurs interrelations. C'est pourquoi on subdivise habituellement le composant G en un composant A (addition cumulée des effets de chaque gène) et un composant D (effet résultant des interactions dominantes entre gènes). L'action de l'environnement se subdivise en une composante E strictement environnementale et une composante I décrivant les interactions entre gènes et environnement :

$$P = A + D + E + I.$$

Ces différentes composantes ne sont pas déterminées pour un seul individu mais pour des populations en évaluant les variances de chacune de ces populations : VP = VA + VD + VE + VI.

En référence à la notion présentée plus haut on définit l'héritabilité au sens large d'un trait, soit H^2 comme la part de la variance P (la variation phénotypique totale) due à la variance G (VA + VD). Fréquemment le terme d'héritabilité (h^2) fait référence à la seule variance additive des gènes (VA) intervenant dans la formation d'un trait. C'est en ce sens qu'on peut dire d'un gène ou d'un locus contenant une instruction génique et influençant le trait d'intérêt qu'il possède une faible ou une forte héritabilité.

Dans l'estimation de ces variances il est essentiel de tenir compte des relations exactes de parenté. Un père et son fils partage 50 % de leurs gènes tout comme la mère et le fils. Ce n'est pas par contre le cas du père et de la mère dont les ascendants sont généralement différents. De façon similaire, deux sœurs ont en commun 50 % de leurs gènes alors que deux demi-sœurs n'en partagent que 25 %. Il est nécessaire de tenir compte de ces différences parentales pour estimer la proportion de variance phénotypique totale attribuable à chaque composante de VP.

S'appuyant sur l'analyse des génomes entiers, la génétique quantitative fait référence aujourd'hui à la notion de QTL. Déterminer un QTL nécessite donc de combiner avec le plus de précision possible des mesures phénotypiques, généalogiques et génétiques sur un grand nombre d'individus. Son domaine d'action principal concerne donc les traits variables qui sont déterminés par un nombre généralement assez élevé de gènes (ou d'instructions géniques).

Paramètres statistiques fréquemment utilisés en génétique quantitative

Une classe inclut tous les individus présentant le même phénotype pour un caractère précis. Dans une population, la moyenne $\bar{x}$ pour un caractère correspond à la somme de toutes les valeurs de ce caractère, soit (xi) divisée par le nombre de mesures effectuées (N) :

$$\bar{x} = \frac{1}{N}\sum xi$$

À la valeur moyenne d'un caractère est attribuée la valeur 0 en abscisse. Pour un individu donné, la valeur x de l'écart à la moyenne est positive ou négative selon la valeur phénotypique de cet individu par rapport à la classe moyenne.

La fréquence relative f_i exprime pour un caractère variable la proportion des individus (n_i) appartenant à une classe (x_a) de ce caractère à tous les individus N analysés :



$$f_i = \frac{n_i}{N}.$$

La variance (s^2) exprime l'amplitude de la dispersion des différentes classes d'individus (x_i ou classes de caractères) par rapport à la classe centrale, $\bar{x}$:

$$s^2 = \frac{\sum (x_i - \bar{x})}{N}.$$

La variance s'exprime en unités au carré. L'écart-type (s) de la distribution correspond à la racine carré de la variance.

La covariance permet d'établir si chez un individu deux caractères, par exemple X et Y évoluent dans le même sens. On a :

$$\mathrm{cov}_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N}.$$

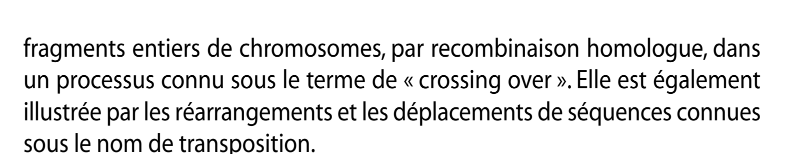
Le coefficient de régression (b_{xy}) permet de déterminer si la valeur pour un caractère (Y) dépend de celle du caractère (X). C'est la pente de la droite Y = *f* (X), soit :

$$b_{xy} = \frac{\mathrm{cov}\ xy}{s^2(x)}.$$



POINTS CLEFS

- Pour survivre et se reproduire, les organismes vivants doivent répliquer fidèlement leur ADN et le protéger des détériorations physico-chimiques.
- Si l'ADN était absolument invariant, c'est-à-dire répliqué avec une fidélité absolue et donc dépourvu totalement de mutations, il n'y aurait pas d'évolution des espèces vivantes.
- Un taux élevé de mutations peut permettre à certains organismes d'évoluer et donc de s'adapter plus rapidement à leur environnement.
- Les mutations, qu'elles soient spontanées ou provoquées par des détériorations liées au contexte environnemental, sont principalement dues à des erreurs de duplication de l'ADN, non corrigées.
- Contrairement aux idées prédominantes en vogue après la découverte de sa structure, insistant sur le caractère invariant de l'ADN, on considère aujourd'hui la recombinaison et plus largement la mobilité et la plasticité de la séquence, comme des caractéristiques courantes et indispensables au fonctionnement et à la reproduction des êtres vivants. Cette réalité se manifeste particulièrement à la méïose, notamment lors de l'échange de


fragments entiers de chromosomes, par recombinaison homologue, dans un processus connu sous le terme de « crossing over ». Elle est également illustrée par les réarrangements et les déplacements de séquences connues sous le nom de transposition.

- Qu'elles soient géniques ou chromosomiques, les mutations ont de multiples formes ponctuelles (transitions, transversions) ou plus larges (amplifications, inversions, délétions, translocations, transpositions, recombinaisons diverses). Elles peuvent entraîner des pertes d'hétérozygotie. Leurs effets se manifestent généralement par des modifications phénotypiques mais pas toujours. Elles peuvent être favorables, défavorables ou neutres suivant le contexte environnemental et la sélection qui prévaut.
- La sélection est un processus biologique complexe à la base de l'évolution des espèces. Il se déroule dans la nature (sélection naturelle) mais peut aussi être dirigé à son profit par l'homme (sélection artificielle).
- La sélection naturelle aboutit à ce que les caractéristiques (traits phénotypiques) favorables, transmises entre individus capables de se reproduire deviendront de plus en plus fréquentes chez les descendants au cours des générations successives. Si les caractéristiques ainsi sélectionnées ont une base génétique, le génotype correspondant deviendra également de plus en plus fréquent dans les générations suivantes.
- La sélection naturelle peut se manifester à divers stades du cycle vital d'un organisme ayant une reproduction sexuée.
- La sélection naturelle de leurs caractéristiques phénotypiques assure aux individus porteurs une excellente adaptation au milieu dans lequel ils vivent (niche écologique). Le maintien d'une telle sélection basée sur un processus interactif entre génotype et milieu peut aboutir finalement à l'émergence d'une nouvelle espèce. Cette idée simple repose sur le concept d'adaptation et demeure aujourd'hui encore à la base de la théorie de l'évolution des espèces vivantes, même si celle-ci s'enrichit depuis les années 1950, de nombreuses connaissances cellulaires et moléculaires.
- La sélection est un processus qui en éliminant les mutations défavorables à une adaptation environnementale réduit la variabilité génétique apportée par les mutations. Dans le même temps, de nouvelles mutations, aléatoires, se produisent qui réalimentent la variabilité génétique au sein d'une population. Une balance, un équilibre s'établit entre les deux processus (vitesse d'apparition des mutations et élimination des mutations défavorables). De cette balance résulte en quelque sorte la force propulsive de l'évolution.
- Le jeu complexe qui s'établit entre phénotype, génotype et milieu est parfois représenté par la relation : génotype + environnement + variation aléatoire $\rightarrow$ phénotype.



- Quand les locus de deux gènes (allèles) sont liés ou très proches l'un de l'autre sur le même chromosome, on dit qu'existe une liaison génétique.
- Le déséquilibre de liaison décrit la situation qui dans une population résulte de certaines combinaisons d'allèles ou de marqueurs génétiques dont la fréquence diffère de celle qu'on attend d'une formation aléatoire d'haplotypes. Autrement dit, les associations non aléatoires entre gènes (allèles) à différents loci sont mesurées par le degré de déséquilibre de liaison.
- Les instructions génétiques (l'inné) ne sont pas les seules causes des caractéristiques biologiques d'un individu puisque l'on sait que des facteurs environnementaux (l'acquis) peuvent aussi tout au long de la vie les moduler. Mais l'hérédité c'est la part génétique au sens strict. La notion d'héritabilité fait référence précisément à l'estimation de cette part génétique en ce qu'elle peut expliquer les différences phénotypiques constatées entre individus d'une même population.
- La génétique quantitative étudie les caractéristiques biologiques (les traits phénotypiques) qui sont mesurables et quantifiables, en essayant de relier la série des valeurs constatées et leur distribution (la variabilité des phénotypes) aux mécanismes sous-jacents tels qu'ils sont admis par une majorité de généticiens se référant aux bases théoriques actuelles de leur discipline.

QCM - QROC

2.1 Qu'est-ce que l'hérédité ?

a) Le processus de transmission des caractéristiques biologiques d'un parent à sa descendance.

b) Ce qui peut expliquer les différences phénotypiques constatées entre individus d'une même population.

c) Ce qui dans la formation d'un organisme relève des parts respectives des gènes et de l'environnement.

d) Ce qui appartient à l'inné et ce qui appartient à l'acquis.

2.2 Qu'est-ce que la sélection ?

a) Un processus qui aboutit à ce que les caractéristiques (traits phénotypiques) favorables, transmises entre individus capables de se reproduire deviennent de plus en plus fréquentes chez les descendants au cours des générations successives.

b) Un processus qui aboutit à ce que la fréquence d'un allèle augmente parce que son adaptabilité sera plus grande que celle des autres allèles.

c) Un processus qui agit sur le potentiel de variation du génome et suivant les circonstances environnementales et l'échelle de temps considérée, peut ou non appauvrir la variabilité intrinsèque du génome.

2.3 Qu'entend-on par l'expression « déséquilibre de liaison » :

a) une distribution non aléatoire de marqueurs génétiques ?

b) une distribution aléatoire de marqueurs génétiques ?

c) une absence de marqueurs génétiques ?

d) une abondance de marqueurs génétiques ?

2.4 Citer deux aspects du processus reproductif d'une espèce où peut s'exercer la sélection.

2.5 Citer deux exemples de mutations à effet favorable chez l'homme.

2.6 Citer deux détériorations chimiques provoquant des mutations.

2.7 Pourquoi une mutation provoquant un changement de cadre de lecture entraîne-t-elle plus fréquemment qu'une mutation « faux sens », la perte de fonction de la protéine correspondante ?

2.8 Chez un couple sans antécédents familiaux naît un enfant de taille normale et un autre atteint de nanisme achondroplasique (arrêt du développement de certains os), un phénotype autosomal dominant, très étroitement associé à un génotype mutant. Comment expliquer la naissance de cet enfant nain ? D'autres enfants de ce couple peuvent-ils être atteints de ce type de nanisme.

RÉPONSES

2.1 a)

2.2 a), b), c)

2.3 a)

2.4 Capacité à se reproduire (ne pas disparaître avant d'être capable de se reproduire), Compétition sexuelle (compétition entre adultes pour se reproduire), Nombre de descendants (et durée de la période d'aptitude à la reproduction), Fécondité croisée (fécondité relative mâle et femelle), Viabilité et sélection des gamètes (suivant les événements de recombinaison méïotique), Compatibilité des gamètes (plus ou moins grande compatibilité sperme-ovule).



2.5 Une mutation ponctuelle de la globine (un acide glutamique est substitué par la valine) responsable de l'anémie falciforme (l'agrégation de l'hémoglobine déforme les hématies qui adoptent une forme en faucille et circulent mal). Les porteurs de cette mutation résistent à la malaria (paludisme), l'agent infectieux de cette maladie (*Plasmodium*) étant mal transporté.

Une délétion de 32 bases dans le gène codant un récepteur de cytokines (CCR5), exprimé préférentiellement dans les lymphocytes T, les macrophages et les cellules dendritiques. Ce récepteur, utilisé par le VIH (virus du sida) comme corécepteur pour coloniser les lymphocytes T, est non fonctionnel sous sa forme mutée (CCR5-Δ32) et confère aux porteurs homozygotes une résistance à l'infection par le VIH.

2.6 L'hydrolyse spontanée de la liaison N-glycosidique unissant les bases puriques au désoxyribose crée des pseudo-nucléotides sans guanine ou adénine, dits apuriniques qui bloquent la réplication. Si malgré tout une base est insérée en face de tels pseudo-nucléotides, il y a une chance sur quatre qu'elle soit erronée créant ainsi une mutation. L'hydrolyse spontanée des groupes NH_2 de la cytosine ou de la 5-méthylcytosine donne naissance respectivement à l'uracile et à la thymine. Elles s'apparieront alors après un cycle de réplication avec l'adénine et non plus avec la guanine créant ainsi une mutation ponctuelle par transition. La 7,8-dihydro-8-oxoguanine (oxoG) formée par certains radicaux libres oxygénés, s'apparie aussi bien avec la cytosine que l'adénine, provoquant dans ce dernier cas une des mutations par transversion les plus fréquemment rencontrées à l'origine de certains cancers humains.

2.7 Le changement de cadre de lecture résulte au minimum de la délétion ou de l'addition d'une base (jamais de trois bases ou d'un multiple de trois). Du site de la mutation jusqu'à l'extrémité de la protéine, la séquence en amino-acides s'en trouve modifiée, avec très souvent aussi la possibilité d'apparition d'un codon stop prématuré et la formation d'une protéine tronquée. La mutation « faux sens » change uniquement le sens du codon concerné et donc un seul amino acide.

2.8 L'allèle responsable de ce nanisme doit être apparu spontanément dans la lignée germinale de l'un des parents. La survenue d'un autre enfant atteint de nanisme dépend du stade auquel la mutation est intervenue (cellule souche de la lignée germinale ou gamètes). Si les cellules souches (2 *n* chromosomes) de la lignée germinale de l'un des parents ont muté spontanément, la moitié des gamètes (*n* chromosomes) seront susceptibles d'être porteurs et la probabilité de survenue

d'un autre enfant nain sera de 1/2. Elle sera nulle, si l'événement de mutation initial est intervenu après la méiose ayant produit les gamètes porteurs. Elle sera comprise entre 0 et 0,5 selon la proportion de gamètes atteints aux autres stades possibles de mutation.





Structure des gènes et des génomes

PLAN

OBJECTIFS

- Analyser dans ses principaux aspects la structure en double hélice de l'ADN
- Analyser dans leurs principaux aspects la structure et la dynamique des nucléosomes et de la chromatine
- Répertorier les principaux aspects de la structure et de l'organisation des différents génomes

3.1 LA STRUCTURE DE L'ADN

Nucléotides et polynucléotides

La structure d'un polynucléotide (polymère de nucléotides) repose sur la formation d'un nombre parfois élevé (plusieurs millions) de liaisons **phosphodiester** entre les nucléotides constitutifs (voir le Mini Manuel *Biologie moléculaire*). Chaque liaison phosphodiester s'établit entre le groupe OH du carbone 3' d'un pentose et le groupe OH porté par le carbone 5' du pentose suivant. C'est la liaison phosphodiester 3'-5'. Il en résulte que l'orientation du polynucléotide se trouve définie par la position du groupe phosphate libre du premier nucléotide, généralement le phosphate porté par le carbone 5', et par le groupe OH libre porté par le dernier nucléotide, généralement le phosphate porté par le carbone 3'. On dit ainsi que le polynucléotide est orienté dans le sens 5' vers 3'.

L'ADN est un acide nucléique **bicaténaire**, c'est-à-dire constitué de deux brins polynucléotidiques associés par des **liaisons hydrogène** entre les bases puriques et pyrimidiques constitutives des nucléotides. Les liaisons hydrogène s'établissent toujours entre une purine de l'un des brins et une pyrimidine de l'autre brin. Dans cet appariement complémentaire, l'adénine (A) est donc toujours associée à la thymine (T) par deux liaisons hydrogène et la guanine (G) interagit toujours avec la cytosine (C) grâce à trois liaisons hydrogènes *(Fig. 3.1)*. Ainsi, dans



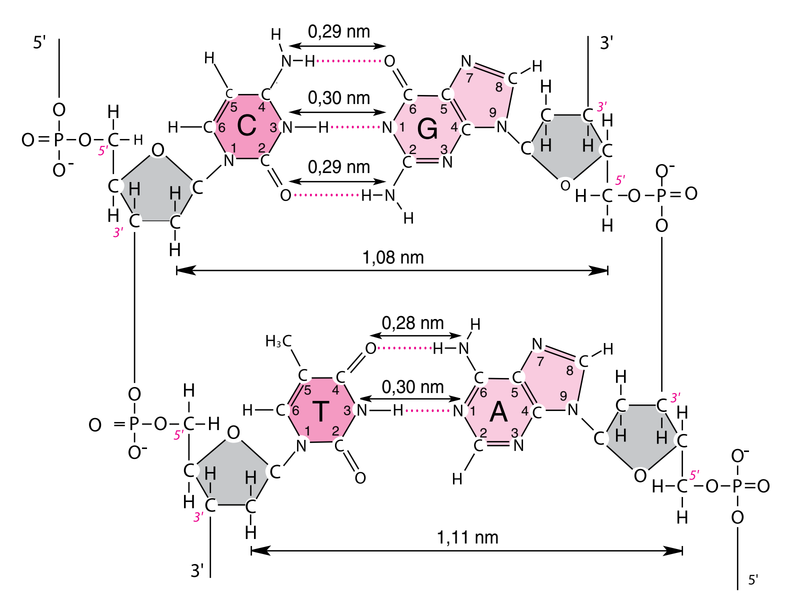
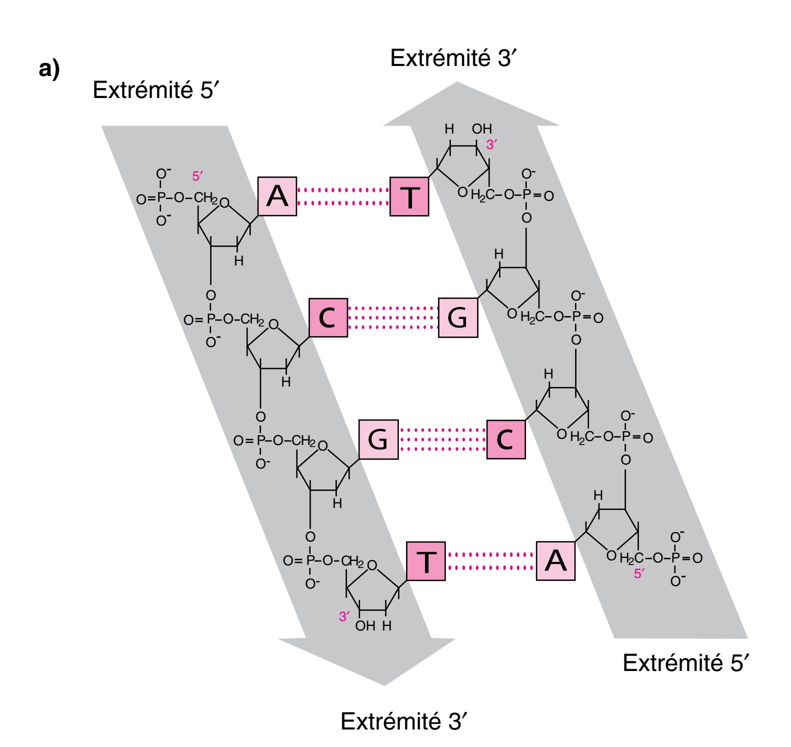


Figure 3.1 Appariements complémentaires des bases G-C et A-T.

Tableau 3.1 Composition en bases de l'ADN de différentes origines.

Origine de l'ADN	Bases (%)				(A + T)/(G + C)	(A + G)/(T + C)
	A	G	C	T		
Bactériophage T7	26,0	23,8	23,6	26,6	1,11	0,99
Escherichia coli B	23,8	26,8	26,6	23,1	0,88	1,01
Neurospora	23,0	27,1	26,6	23,3	0,86	1,00
Drosophile	30,7	19,6	20,2	29,5	1,51	1,01
Saumon	28,0	22,0	20,0	27,8	1,33	1,05
Poule	28,0	22,0	21,6	28,4	1,29	1,00
Rat	28,6	21,4	21,6	28,4	1,33	1,00
Vache	27,3	22,5	22,5	27,7	1,26	0,99
Homme	29,3	20,7	20,0	30,0	1,46	1,00



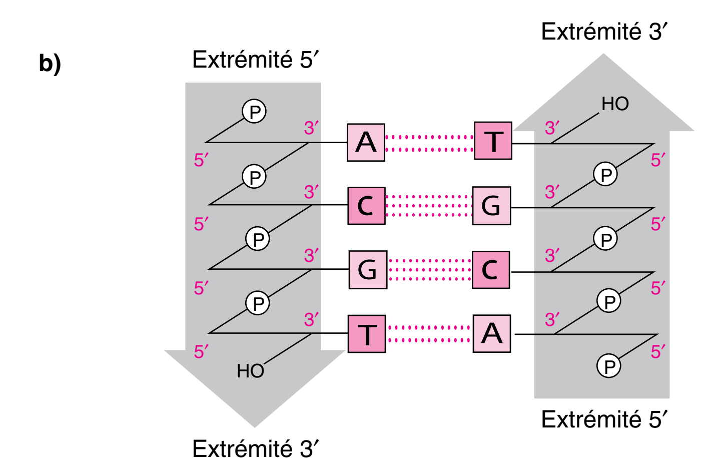


Figure 3.2 Structure des appariements complémentaires de deux brins anti-parallèles d'ADN. **a)** représentation développée ; **b)** représentation simplifiée.



l'ADN, la quantité des purines (A + G) est égale à celle des pyrimidines (C + T). Par contre le rapport A + T sur G + C varie selon l'origine de l'ADN *(Tableau 3.1* et voir le Mini Manuel *Biologie moléculaire, Tableau 1.1)*. Le principe d'**appariement complémentaire** des bases a une grande valeur pratique. Il permet au généticien de déterminer la séquence nucléotidique d'un brin d'ADN à partir de celle du brin complémentaire, qu'il s'agisse d'un brin ADN ou d'un transcrit ARN. L'orientation d'un brin polynucléotidique est définie par la présence des groupes chimiques 5'-phosphate et 3'-OH portés par les carbones 5' et 3' des deux désoxyriboses extrêmes *(Fig. 3.2)*. Dans la **double hélice d'ADN**, les deux brins complémentaires sont orientés en sens inverse. On dit qu'ils sont **antiparallèles**.

La double hélice

L'association des deux brins par des liaisons hydrogènes provoque une contrainte physique qui impose à l'ensemble l'adoption d'une structure en double hélice *(Fig. 3.3)*. Dans le modèle original, proposé par James Watson et Francis Crick en 1953, la double hélice a un diamètre de 2 nm et présente deux périodicités. La première de 0,34 nm correspond à la distance séparant deux plateaux de bases appariées. La seconde de 3,4 nm correspond au pas de l'hélice qui couvre une longueur de 10,5 plateaux de bases *(Fig. 3.3)*. À la surface de la molécule

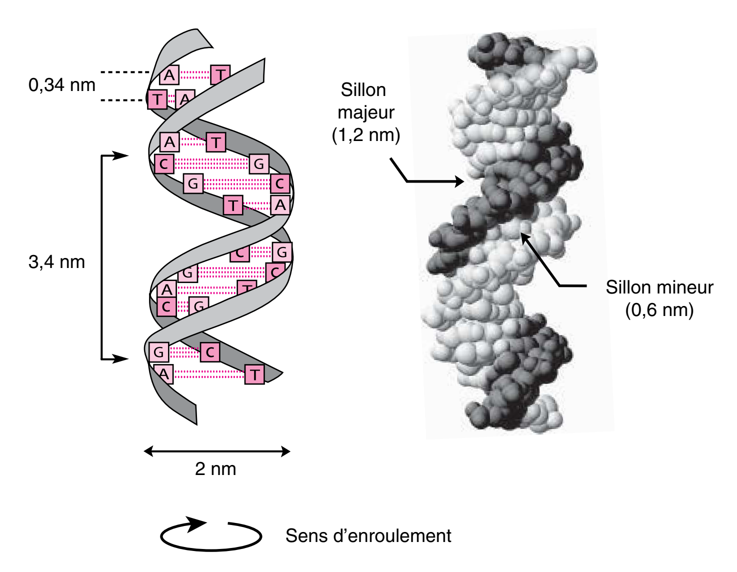


Figure 3.3 Représentation d'un fragment d'ADN en double hélice.

À gauche, avec les plateaux de bases ; à droite, les atomes sont représentés avec leurs rayons de Van der Waals.

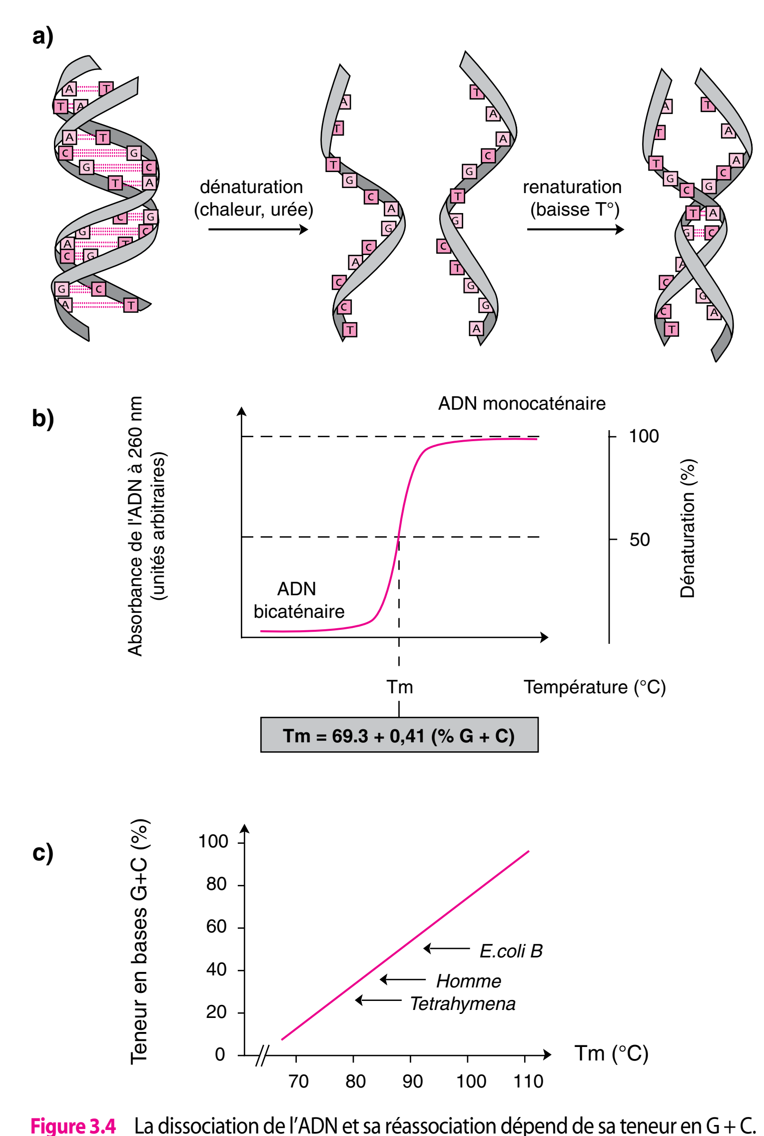


Figure 3.4 La dissociation de l'ADN et sa réassociation dépend de sa teneur en G + C.

a) Représentation simplifiée du processus de dissociation-réassociation de l'ADN sous l'action d'agents dénaturants ; **b)** représentation simplifiée de la courbe sigmoïde, témoin du processus de dissociation. La formule établit la relation entre la teneur en G + C de l'ADN et la température de fusion de l'ADN ; **c)** droite reliant la teneur en G + C et la Tm (Température de fusion) de divers organismes.



d'ADN, les deux brins torsadés ménagent deux **sillons hélicoïdaux** de largeur différente, un sillon majeur ou grand sillon et un sillon mineur ou petit sillon *(Fig. 3.3)*. Ces deux sillons, tout particulièrement le grand sillon, exposent de nombreux groupes chimiques avec lesquels interagissent dans les conditions physiologiques de la cellule des protéines dites de liaison à l'ADN. Ces protéines assument des rôles majeurs de régulation de l'expression de l'information génétique portée par la séquence en bases proprement dite. Sous sa forme bicaténaire (en double hélice) l'ADN absorbe modérément la lumière UV. Après dissociation des deux brins, le démasquage des bases qui en résulte, provoque une absorption plus marquée de la lumière UV. L'enregistrement du phénomène à 260 nm fournit une courbe en forme de sigmoïde, typique du phénomène *(Fig. 3.4)*. La dissociation thermique de l'ADN et sa réassociation par refroidissement est une technique de base utilisée dans la réaction de polymérisation en chaîne (PCR).

Sous la forme linéaire souvent observée dans la plupart des cellules, les extrémités de l'ADN sont libres et les contraintes physiques introduites par ses divers ligands sont, pour cette raison, facilement ajustées. L'ADN peut cependant exister sous une forme circulaire, c'est-à-dire sans extrémités libres. C'est le cas de l'ADN bactérien, de beaucoup d'ADN viraux et de l'ADN mitochondrial *(Tableau 3.2)*. C'est aussi le cas des **plasmides**, petits chromosomes bactériens surnuméraires capables de s'autorépliquer de façon autonome. L'ADN circulaire subit des contraintes physiques qui ne sont réduites pour stabiliser l'ensemble que par un jeu de surenroulements parfois nombreux. Quel que soit l'ADN, l'orientation des gènes dépend pour chaque gène de l'orientation du brin effectivement transcrit (orientation 3' → 5'). Les deux brins peuvent être porteurs de l'information transcrite *(Fig. 3.5)*. Des chevauchements existent parfois entre gènes dont l'orientation est opposée.

Dans l'ADN de très grande longueur des cellules eucaryotes, malgré la linéarité de la molécule, les **contraintes topologiques** demeurent au sein de la chromatine. Il est donc important que ces contraintes soient réduites afin que les processus fondamentaux que sont la réplication et la transcription de l'ADN soient possibles. Une classe spéciale d'enzymes, nommées **topoisomérases**, peut introduire des cassures spécifiques dans l'un ou l'autre brin ou même les deux. Ces cassures réduisent les contraintes, relâchent les surenroulements et rendent possibles les processus physiologiques. Ceci est particulièrement important lors de la séparation des deux molécules d'ADN résultant de la réplication, qui sont encore liées l'une à l'autre avant leur adressage aux deux cellules produites par la division cellulaire.

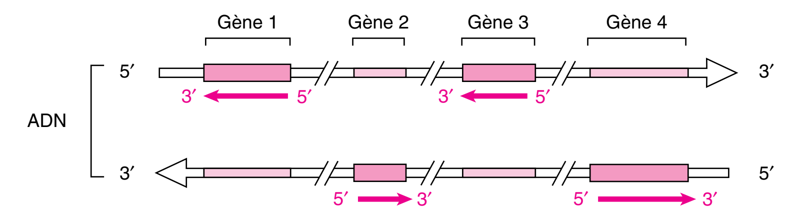


Figure 3.5 Orientation des gènes suivant que la séquence codante se trouve sur l'un ou l'autre brin de la double hélice.

TABLEAU 3.2 GÉNOMES DE DIVERS ORGANISMES : STRUCTURE, TAILLE ET ESTIMATION DU NOMBRE DE GÈNES.

Génome	Groupe	Structure du génome	Taille (kb)	Nombre de gènes
Eucaryotes :				
Saccharomyces cerevisae	Levures		13 500 (L)	6 000
Caenorhabditis elegans	Nématode		100 000 (L)	13 500
Arabidopsis thaliana	Plante		120 000 (L)	25 000
Homo sapiens	Humain		3.10^6 (L)	28 000
Procaryotes :				
Escherichia coli	Bactérie		4 700 (C)	4 000
Haemophilus influenza	Bactérie		1 830 (L)	1 703
Methanococcus jannaschii	Bactérie		1 660 (L)	1 738
Virus :				
Virus T4	Virus bactérien		172 (L/C)	300
HCMV (virus de l'herpès)	Virus eucaryote		229 (L)	200
Organites				
Mitochondrie de *S. cerevisae*	Levure		78 (C)	34
Mitochondrie de *H. sapiens*	Humain		17 (C)	37
Chloroplaste	Plante		121 (C)	136

Séquences répétées
Introns
Exons

L : Linéaire
C : Circulaire



3.2 LA STRUCTURE ET LA FONCTION DES CHROMOSOMES EUCARYOTES

La structure du nucléosome

Dans les cellules eucaryotes, à la différence des procaryotes, l'ADN est très fréquemment associé à des protéines basiques, les **histones**. Le résultat de l'assemblage régulier entre ADN et histones est désigné par le terme de **nucléosome** *(Fig. 3.6)*.

La formation des nucléosomes est la première étape d'un processus qui assure l'empaquetage de l'ADN en structures compactes réduisant énormément le volume occupé par la molécule. L'ensemble est désigné sous le terme de **chromatine**.

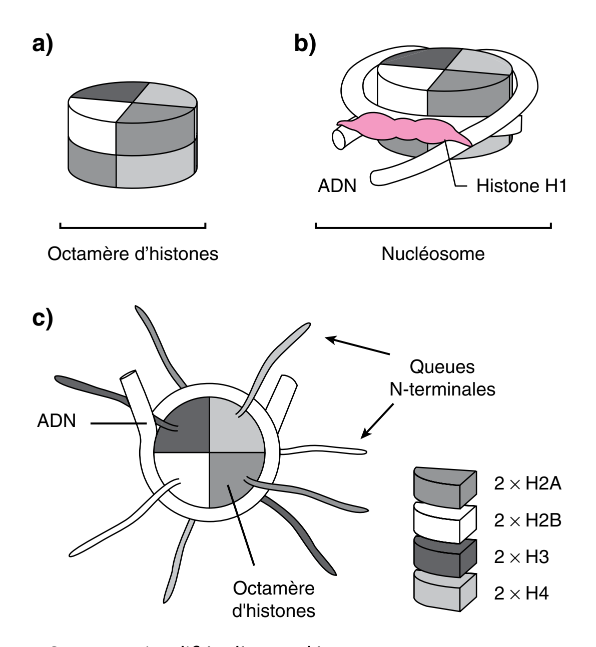


Figure 3.6 Structure simplifiée d'un nucléosome.

a) l'octamère d'histones ; **b)** l'enroulement de l'ADN autour de l'octamère et sa solidification par l'histone H1 ; **c)** protubérance vers l'extérieur des queues N-terminales d'histones de l'octamère.

Chacune des classes d'histones est constituée de petites protéines riches en amino-acides chargés positivement (lysine et arginine). Leurs séquences en amino-acides sont restées remarquablement identiques au cours de l'évolution. Les **histones H2A**, **H2B**, **H3** et **H4** (11 à 15 kd) sont en quantité égale et chacune en double exemplaire, elles forment un **octamère** d'histones *(Fig. 3.6 a)* constituant un « cœur » protéique

discoïde de 8 protéines autour duquel l'**ADN nucléosomal** d'une taille de 147 paires de bases est enroulé *(Fig. 3.6 b)*. L'histone **H1** (20 kd) ne fait pas partie du « cœur ». Elle est attachée d'une part, à l'ADN de liaison situé entre les nucléosomes et, d'autre part, à l'ADN nucléosomal ce qui a pour effet de solidifier l'ensemble *(Fig. 3.6* et *3.7)*.

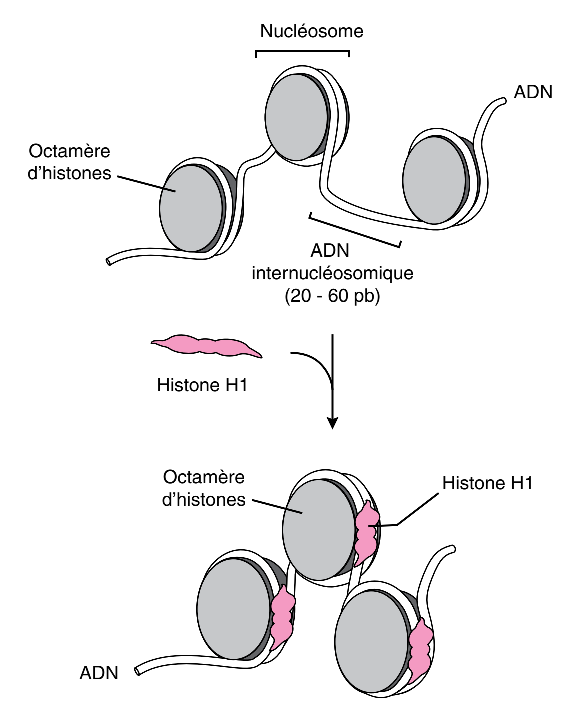


Figure 3.7 Représentation simplifiée d'un enchaînement de trois nucléosomes montrant le rôle de l'histone H1 dans le resserrement de l'ADN autour de l'octamère d'histones.

De nombreux contacts existent entre les histones de l'octamère et l'ADN nucléosomal. Ils sont indépendants de la séquence en bases de l'ADN, ce qui explique que la presque totalité de l'ADN des cellules eucaryotes est organisée en nucléosomes. Les histones du cœur possèdent des extrémités libres du côté N, désignées par le terme de « queues ». Ces queues émergent du nucléosome en des positions très précises *(Fig. 3.6 c)*. Elles facilitent la formation des structures d'ordre supérieur résultant des surenroulements de nucléosomes donnant la structure finale de la chromatine *(Fig. 3.8)*. Certains amino-acides des **queues d'histones** sont parfois modifiés par des groupes acétyle et méthyle (lysine, arginine) ou phosphate (sérine). Ces modifications

Double hélice d'ADN
(2 nm)

11 nm

Nucléosomes formant
une structure en
« collier de perles »

Fibre compactée de
chromatine de 30 nm
de diamètre.
Structure en solénoïde
(6 nucléosomes/tour)

Filament de 300 nm
de diamètre.
Domaine en boucles

Matrice protéique

Filament superenroulé
de 700 nm de diamètre

Chromosome
métaphasique condensé
(1 400 nm)

Figure 3.8 Modèle hypothétique de l'organisation et de la compaction de la chromatine.



facilitent ou altèrent l'accessibilité de la chromatine. Elles constitueraient un « code épigénétique » qui pourrait être lu par les protéines impliquées dans la régulation de l'expression des gènes et dans les diverses autres fonctions de l'ADN. Les nucléosomes sont des structures dynamiques, constamment modifiées par de nombreux complexes protéiques de remodelage capables de les déplacer le long de la molécule d'ADN, de les transférer d'une molécule d'ADN à une autre ou de les modifier afin d'accroître l'accessibilité de l'ADN.

La structure et le remodelage de la chromatine

La présence de l'histone H1 solidifie et resserre la structure du nucléosome. L'ADN nucléosomal représente alors 160 pb. L'histone H1 favorise aussi la formation de structure d'ordre supérieur. Les nucléosomes s'organisent alors en une **fibre de 30 nm** de diamètre qui est en fait une **super-hélice** comportant 6 nucléosomes par tour. Ceci ne suffit pas cependant pour empaqueter les 1 à 2 mètres d'ADN dans un noyau cellulaire de 10^{-5} mètre environ de diamètre. Des repliements en forme de **boucles** de la fibre de 30 nm sont donc nécessaires. Les boucles seraient maintenues dans un ensemble compact par leur association à un support protéique, une sorte d'échafaudage protéique nucléaire à vrai dire encore un peu mystérieux *(Fig. 3.8)*. Lorsque les structures d'ordre supérieur prédominent, cela a pour effet de réduire fortement l'accessibilité de l'ADN aux enzymes et autres protéines nécessaires à la transcription. À l'inverse, le remodelage strictement contrôlé de la chromatine par des complexes protéiques spécialisés est une condition absolue de l'expression génique.

La structure des chromosomes au cours du cycle cellulaire

Dans les cellules eucaryotes, le **chromosome** est le stade supérieur et ultime d'organisation de l'ADN (voir *Fig. 3.8*). Le nombre de chromosomes et leur composition sont des caractéristiques de chaque organisme.

Le séquençage complet d'un certain nombre de génomes a montré que la quantité d'ADN effectivement utilisé pour coder les protéines ou les ARN varie énormément d'un organisme à un autre. Chez les bactéries, la presque totalité de l'ADN code des protéines. Chez les mammifères, seule une petite portion (moins de 10 %) code réellement les protéines ou les divers ARN, le reste ou **ADN intergénique**, est constitué largement de séquences répétées (souvent des dinucléotides), appelées **microsatellites** (3 %) ou composées de grands éléments de **séquences transposables** (plus de 50 %). Beaucoup d'interrogations demeurent encore aujourd'hui sur les fonctions exactes de l'ADN



intergénique, dont l'apparente inutilité semble contredite par sa nécessaire et exacte duplication à chaque division cellulaire.

Pour être maintenu comme tels au cours de la division cellulaire, chaque chromosome doit posséder un **centromère**, des **télomères** et plusieurs **origines de réplication**. Les origines de réplication sont les nombreux sites (3000 par chromosome en moyenne) où s'assemblent les divers composants protéiques, dont les **ADN polymérases**, nécessaires au démarrage de la réplication (duplication de l'ADN) opération qui précède la division cellulaire. Les centromères sont des séquences d'une taille supérieure à 40 kb au contact desquelles se construit le **kinétochore**, un complexe protéique nécessaire à la séparation et l'adressage corrects des deux chromosomes-fils aux deux cellules filles.

Il est essentiel qu'un seul centromère soit présent par chromosome. Dans le cas contraire, il en résulte des cassures, des pertes ou des duplications de chromosomes dont les conséquences sont généralement graves.

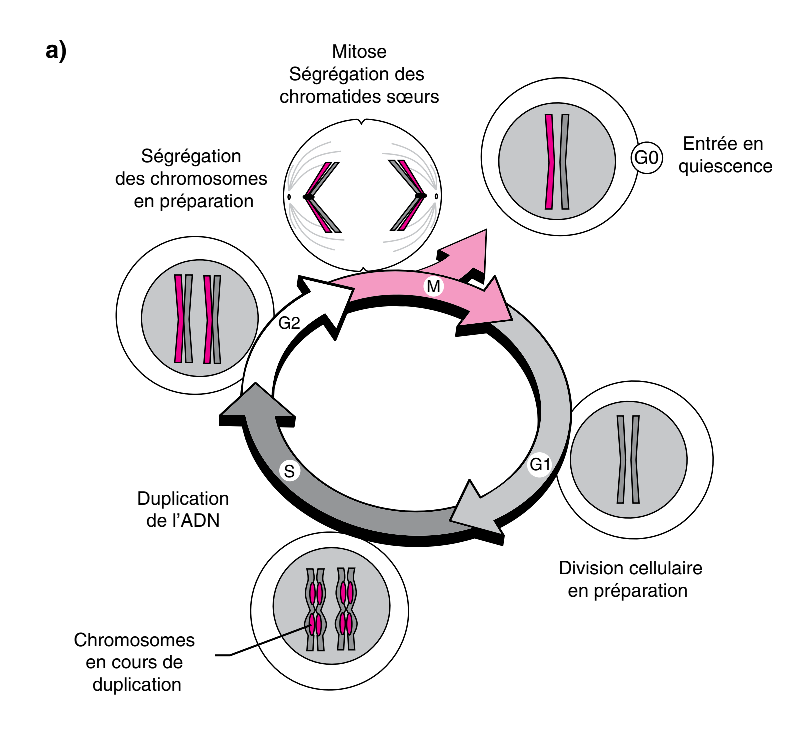
Les télomères sont des séquences particulières localisées aux deux extrémités du chromosome. Elles sont reconnues par des protéines spécifiques qui protègent le chromosome des événements de recombinaison ou de dégradation typiques des extrémités ADN libres. À l'endroit des télomères se trouve une origine de réplication particulière reconnue par une ADN polymérase spéciale, la **télomérase**, qui réplique un certain nombre de fois la séquence correspondante.

La duplication du chromosome puis sa ségrégation s'opère dans des phases distinctes du **cycle cellulaire** (division cellulaire ou mitose). On distingue dans l'ordre 4 phases : **G1**, **S**, **G2** et **M** *(Fig. 3.9)*.

Au cours de la phase S s'opère la synthèse de l'ADN conduisant à la duplication des chromosomes en deux **chromatides** sœurs maintenues ensemble par un complexe protéique appelé **cohésine**, jusqu'à la ségrégation.

Celle-ci s'opère durant la phase M (pour mitose). Chaque chromatide s'attache grâce au kinétochore à un fuseau mitotique fait de protéines fibreuses appelées **microtubules** réunies à deux pôles opposés organisés, nommés **centrosomes**. La cohésine en se désagrégeant favorise la séparation des deux chromatides sœurs. L'ensemble des opérations est soigneusement coordonné afin que la séparation et l'adressage assurent un transfert complet et efficace des informations génétiques portées par les deux chromosomes-fils.

Les phases G1 et G2 (G pour gap) sont des phases d'attente et de préparation des phases suivantes. Des mécanismes de contrôle du cycle opèrent pendant ces phases qui autoriseront ou non sa poursuite jusqu'à son terme.



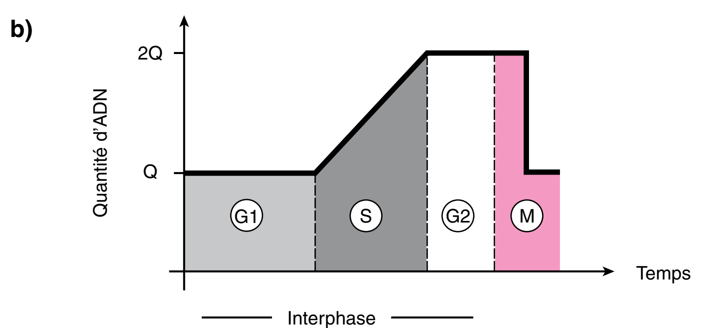


Figure 3.9 Chromosomes et cycle cellulaire.

a) Duplication et ségrégation des chromosomes au cours du cycle cellulaire. **b)** Variation de la quantité d'ADN « Q » au niveau des chromosomes à un instant *t* dans la cellule.

Au cours du processus complet de la division cellulaire, l'état de condensation des chromosomes varie profondément. Durant les phases G1, S et G2 (groupées sous le terme d'**interphase**) il est plus faible que durant la phase M bien que des changements discrets se produisent constamment.



3.3 LA STRUCTURE DES GÉNOMES

Qu'est-ce qu'un génome ?

On désigne par le terme de **génome** (contraction des mots **gèn**e et chromos**ome**) l'ensemble des informations génétiques, codées par l'ADN chez la plupart des organismes et parfois par l'ARN pour certains virus et qui sont héréditairement transmissibles. De façon plus précise, le terme désigne la séquence de l'ADN correspondant à un jeu **haploïde** de chromosomes (*n* chromosomes). Une cellule somatique humaine, **diploïde** (2 *n* chromosomes), contient donc deux génomes (un génome paternel et un génome maternel) alors qu'une cellule germinale (sexuelle), haploïde (ovule ou spermatozoïde), en contient un seul. Le terme génome s'applique aussi bien à l'ADN du noyau cellulaire (**génome nucléaire**) qu'à l'ADN des organites (**génome mitochondrial**, **génome chloroplastique**). Parler du séquençage du génome d'une espèce signifie que l'ordre des nucléotides constituant l'ADN du jeu haploïde des chromosomes non sexuels (autosomes) et sexuels – mâle et femelle – est établi ou en cours de l'être. On parle de **génomique** (contraction des mots *génom*e et informat*ique*) pour désigner l'étude à grande échelle des propriétés générales des génomes grâce à des moyens informatiques de plus en plus puissants. Les premières questions intéressant l'étude des génomes portent sur leur taille, le nombre de gènes qu'ils contiennent (au sens classique du concept), le nombre de chromosomes constitutifs et la manière dont les gènes sont distribués ou organisés.

La taille des génomes

Pour désigner la taille d'un génome on utilise habituellement une unité exprimant le nombre de paires de nucléotides. Pour le millier de paires de nucléotides ou de bases soit 10^3, l'unité est la kilopaire de bases ou **Kpb**, pour le million soit 10^6, l'unité est la mégapaire de bases ou **Mpb**. On pense généralement qu'il existerait une relation progressive entre la taille des génomes, le nombre de gènes qu'ils contiennent et la complexité des groupes d'organismes vivants *(Fig. 3.10* et *Tableau 3.2)*. En vérité, plus la connaissance intime de la structure et du fonctionnement des génomes progresse plus cette relation s'avère improbable. D'une part, la notion même de « complexité » n'est pas très opérationnelle (une plante à fleurs ou une grenouille sont-elles plus complexes qu'un humain ?) ; d'autre part, l'organisation des instructions génétiques au sein des génomes et surtout leur mode d'expression semblent à la fois très diversifiés et variables selon les organismes.

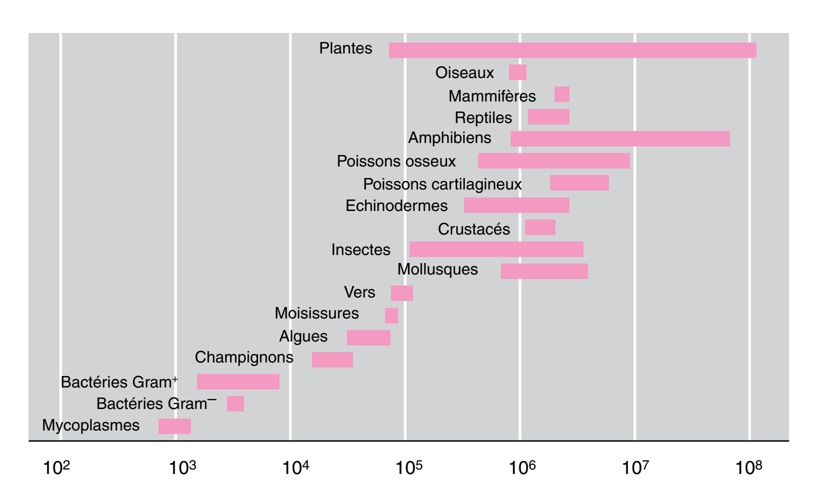


Figure 3.10 Quantité d'ADN présent dans les génomes de différents groupes d'organismes.

Les génomes viraux

Bien que les virus ne soient pas considérés comme des organismes vivants mais comme des parasites moléculaires, ils se reproduisent en infectant des cellules vivantes dans lesquelles ils injectent leur génome constitué par de l'ADN double brins ou parfois simple brin, protégé au sein d'une particule ou capside de nature protéique. Certains virus, comme les rétrovirus auxquels se rattache par exemple le VIH, ont un génome constitué d'ARN, qui se présente soit sous la forme d'un seul brin ou d'un double brin. Quel que soit l'acide nucléique, il est pour certains virus, sous une forme linéaire et pour d'autres, sous une forme circulaire. Dans tous les cas, le génome viral existe au moins dans une des phases de son cycle cellulaire sous la forme d'un ADN double brins.

Les génomes procaryotes

Le génome de la plupart des organismes procaryotes correspond à un seul chromosome composé d'un ADN souvent circulaire et de très peu de protéines associées. Une bactérie en cours de division peut cependant contenir simultanément plusieurs copies de ce chromosome unique. Les gènes bactériens sont très proches les uns des autres avec peu d'**espaces intergéniques**, souvent organisés en unités, les opérons, transcrits en un seul ARN messager (voir le Mini Manuel, *Biologie moléculaire*). L'ADN associé à quelques protéines forme une masse dense, le **nucléoïde** qui se désorganise en un écheveau filamenteux lorsque la cellule est broyée.



Les génomes eucaryotes

Les espèces eucaryotes sont soit diploïdes (les cellules somatiques des animaux et plantes à fleurs possèdent dans leur noyau, deux jeux – 2 *n* – de chromosomes) ou haploïdes (un seul jeu – *n* – de chromosomes par noyau cellulaire chez la plupart des champignons et les algues). Les organismes diploïdes produisent généralement des cellules spécialisées haploïdes (n chromosomes) pour la reproduction (ovule et spermatozoïde chez les animaux) alors que les organismes haploïdes peuvent présenter au cours de la phase sexuée de leur cycle vital des cellules spécialisées à 2 *n* chromosomes. Le nombre *n* de chromosomes présents dans les cellules d'un organisme définit le niveau de ploïdie. Les végétaux présentent souvent des niveaux de **ploïdie** élevés (polyploïdie). Dans une cellule diploïde, les chromosomes s'associent en paires de chromosomes homologues, chaque chromosome possédant une taille et un nombre de gènes identiques aux mêmes positions. Cependant une paire de chromosomes peut porter deux allèles différents d'un même gène (voir plus haut). Le terme de caryotype est utilisé pour désigner le nombre de chromosomes, établi par l'observation microscopique d'une cellule en métaphase.

Au sein d'une espèce, les chromosomes peuvent différer fortement en taille. Chez l'homme *(Tableau 3.3)* les chromosomes 1 et 2 ont une grande taille alors que les chromosomes 21 et 22 sont quatre fois plus

TABLEAU 3.3 PRINCIPALES CARACTÉRISTIQUES DES CHROMOSOMES HUMAINS.

	Grands chromosomes					Chromosomes moyens										Petits chromosomes							Chromosomes sexuels	
Groupe	A			B		C							D			E			F		G		C	G
Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
Longueur relative des chromosomes (%)	8,4	8,0	6,8	6,3	6,1	5,9	5,4	4,9	4.8	4,6	4,6	4,7	3,7	3,6	3,5	3,4	3,3	2,9	2,7	2,6	1,9	2,0	5,1	2,2



petits. Pour mieux les analyser, le cytogénéticien classe les chromosomes d'un organisme suivant leur appartenance à un groupe de taille (de A à G chez l'homme). La position du centromère (voir plus haut) qui apparaît souvent sous la forme d'une constriction chez le chromosome métaphasique, divise le chromosome en deux bras. Le plus court est appelé *p* et le plus long *q*. Le cytogénéticien utilise cette nomenclature pour donner la position d'un gène sur les bras d'un chromosome. On dit d'un chromosome qu'il est **télocentrique** si le centromère est situé à une extrémité, **acrocentrique** s'il est proche d'une extrémité et **métacentrique** s'il est situé vers le milieu.

Les génomes d'organites

À côté des chromosomes nucléaires qui représentent l'essentiel du génome d'un organisme eucaryote, certains organites (essentiellement les mitochondries et les chloroplastes, présents parfois en grand nombre dans une cellule) possèdent des chromosomes sous la forme de petites molécules d'ADN circulaires, d'une taille moyenne avoisinant 10^4 à 10^5 paires de bases, souvent en de nombreuses copies identiques. Pour ces raisons, il arrive que selon les organismes, le nombre de chromosomes d'organites par cellule s'avère très élevé jusqu'à plusieurs centaines parfois. En général, les chromosomes d'organites possèdent des gènes (de l'ordre de 50 à 100), certains avec des introns. Ces gènes sont spécifiques des fonctions assurées par l'organite lui-même. Mais beaucoup des fonctions de l'organite sont spécifiées cependant par des gènes nucléaires sans qu'il y ait de redondance entre les gènes des deux compartiments, mais plutôt une complémentarité. Selon la théorie endosymbiotique relative à l'origine des cellules eucaryotes, mitochondries et chloroplastes seraient d'anciennes bactéries ayant colonisé une cellule hôte elle-même descendante de la première cellule vivante (voir le Mini Manuel *Biologie cellulaire* dans cette série). Au cours de l'évolution, beaucoup des gènes de ces bactéries ancestrales symbiotes ont été transférés dans les chromosomes nucléaires de l'hôte ou parfois tout simplement perdus. Il n'en demeure pas moins que les bénéfices de la symbiose sont suffisamment forts pour que ces organites aient été conservés au cours de l'évolution.

3.4 LE SÉQUENÇAGE DES GÉNOMES

Puisque les instructions qui gouvernent la formation et le fonctionnement des êtres vivants sont contenues dans la séquence de la molécule d'ADN, présente au cœur de toutes les cellules vivantes, il est apparu très tôt, après la découverte du code génétique dans les années 1960,



que la connaissance précise de la séquence spécifique de chaque espèce vivante, constituerait un pas décisif vers le décryptage de questions aussi capitales que celles de l'origine et de l'évolution du vivant. Des pionniers s'illustrèrent dans la mise au point des méthodes de séquençage (le principe du séquençage est présenté dans le Mini Manuel *Biologie moléculaire*), mais il fallut attendre la mise au point de machines performantes et les progrès de l'informatique pour que les premières séquences soient établies.

Fragmenter puis assembler

Les débuts furent modestes. En 1976, la séquence du tout petit génome du bactériophage MS2 fut établie, suivie en 1977 de la séquence du phage ΦX174. C'est à cette occasion que commença à être résolue la question centrale encore présente aujourd'hui pour toute démarche de séquençage. Pour séquencer des molécules d'ADN composées d'un alignement de millions ou de milliards de nucléotides, il s'avère indispensable de fragmenter au préalable la molécule. Mais cette opération fait disparaître l'information de séquence qui est précisément capitale pour l'exploitation des résultats. La solution de cette difficulté est venue de l'utilisation d'algorithmes informatiques capables de combiner les informations de séquence et de reconstituer à partir de petits fragments d'ADN (environ 500 pb), facilement séquençables, la séquence complète d'un génome de grande taille. Cette technique dite du « **shotgun** » connue un grand succès et fût appliquée, à partir des années 1990, pour les très grands génomes comme celui de l'homme, grâce aux progrès de l'automatisation du séquençage et des logiciels d'assemblage informatique. Les fragments assemblés en plus grandes régions génomiques par cette méthode sont dénommés « **contigs** ». Ce terme particulier rappelle qu'il s'agissait bien de reconstituer par l'assemblage de fragments contigus, la continuité de la séquence d'origine. De façon alternative, le développement des techniques de clonage moléculaire ont permis d'obtenir de grandes quantités de chacun des fragments engendrés par coupure enzymatique à l'aide d'endonucléases de restriction ou par cisaillement physique. Des banques furent ainsi constituées à partir desquelles le séquençage des fragments pouvait être organisé et répété autant de fois qu'il le fallait afin de sécuriser les résultats *(Fig. 3.11* et *3.12)*.

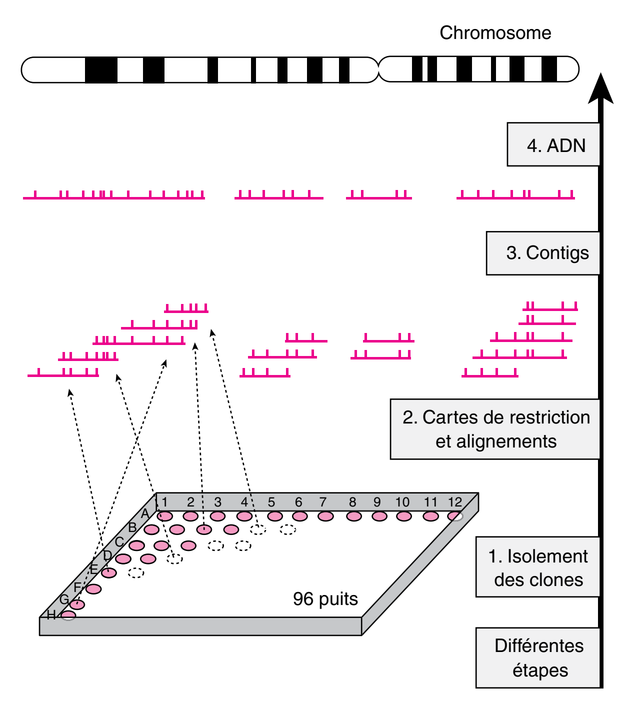


Figure 3.11 Construction et exploitation d'une banque de fragments d'ADN clonés.

L'ADN du chromosome, obtenu lui-même par tri chromosomique en cytométrie de flux est préalablement extrait puis fragmenté par des enzymes de restriction. Plusieurs catégories d'enzymes de restriction sont employées afin d'obtenir plusieurs types de fragments de restriction pour une même séquence. Tous les fragments obtenus sont clonés et mis en culture dans des boîtes spéciales à multiples puits de culture afin de constituer la banque de référence. Chaque clone est ensuite identifié soit par sa carte de restriction, soit par le séquençage de ses extrémités puis les clones sont appariés et ordonnés en contigs en exploitant les chevauchements de séquences qui leur sont communes. Les différents fragments constituant les contigs peuvent être séquencés entièrement et la séquence complète de l'ADN du chromosome sera reconstituée par un assemblage réalisant à l'aide de l'informatique des contigs de contigs. La méthode dite du « shotgun » consiste à fragmenter en morceaux de très petite taille l'ADN chromosomique, à cloner les fragments puis à les séquencer directement, l'assemblage des séquences obtenues s'opérant par le biais de l'informatique pourvu que la puissance des logiciels soit suffisante pour traiter simultanément des millions d'informations unitaires.



Niveau d'organisation de l'ADN

Niveau de résolution cartographique (pb)

Chromosome — 2.10^6 - 2.10^7

Contigs de BACs — $0,5.10^6$ - 2.10^6

(1)

BAC — 4.10^4 - 3.10^5

ER ER ER ER ER

(2)

Sous-clones — 2.10^2 - 4.10^4

Séquence — .......ACGTGTTCCATGTGACACC...... — 1

Figure 3.12 Récapitulation des opérations successives nécessaires à l'obtention de la séquence complète de l'ADN d'un chromosome.

On notera ici le passage par des BACs (chromosome bactérien artificiel contenant des fragment de l'ADN à séquencer d'assez grande taille allant de 4.10^4 à 3.10^5) ce qui facilite la construction de contigs de grande taille et améliore le rendement général de l'opération de séquençage de chaque chromosome.

Le séquençage du génome humain

Ces techniques furent appliquées au séquençage du génome humain dont le projet fut considéré par les autorités internationales comme relevant du franchissement d'une nouvelle frontière de l'humanité. Son coût fut évalué à 3 milliards $ (2,25 milliards €). Ce séquençage a été réalisé concurremment par plusieurs équipes organisées autour



d'un consortium public réunissant plusieurs institutions de différents pays sous la conduite du National Human Genome Research Institut américain dirigé par Francis Collins et par une entreprise privée américaine (Celera Genomics) dirigée par Craig Venter. Les premiers utilisèrent principalement la technique basée sur l'organisation de contigs de **BAC** (Chromosome artificiel bactérien contenant un fragment aléatoire du génome humain, voir *Fig. 3.12*), préservant ainsi au maximum l'ordonnancement de la séquence. Les seconds appliquèrent à ce grand génome la technique de la fragmentation aléatoire dite du « shotgun ». L'essentiel de la séquence (92 %), constituée d'environ 3 milliards de paires de bases, fut obtenue et publiée conjointement en 2003. Les 8 % restants correspondent à des régions hautement répétées (centromères et télomères) ainsi qu'à des régions contenant des familles multigéniques dont on ne sait pas encore si elles pourront réellement un jour être entièrement séquençées.

Le postséquençage

L'un des objectifs du projet de séquençage du génome consistait également à identifier au sein de cet ensemble énorme de données, celles qui définissent les gènes au sens classique du concept (les séquences codant des protéines). Cet aspect du projet reste encore ouvert (*Fig. 3.13*). Le nombre total de gènes n'est pas encore établi avec précision. Il se situerait autour de 25 000, soit beaucoup moins que ce qui était attendu (environ 200 000) sur la base d'un raisonnement, qui s'est révélé faux, liant le nombre de gènes à la complexité supposée de l'organisme étudié.

Un autre objectif du projet consistait à développer les méthodes de séquençage afin de progresser vers l'application de méthodes permettant de déboucher sur une exploitation industrielle assurant des applications systématiques et à faible coût dans le domaine de la santé notamment (voir plus loin et chapitre 5.6).

La séquence établie de l'ADN humain est aujourd'hui déposée dans les banques de données et elle est accessible pour tous par Internet. Un travail intense d'annotation destiné à comprendre comment fonctionnent les instructions contenues dans la séquence a été entrepris. Il s'annonce long et complexe car il dépend en grande partie de connaissances qui restent encore à acquérir (ou à conquérir) sur la nature exacte de ces instructions et sur leur mode d'expression (voir chapitre 1 « Qu'est-ce qu'un gène ? »).

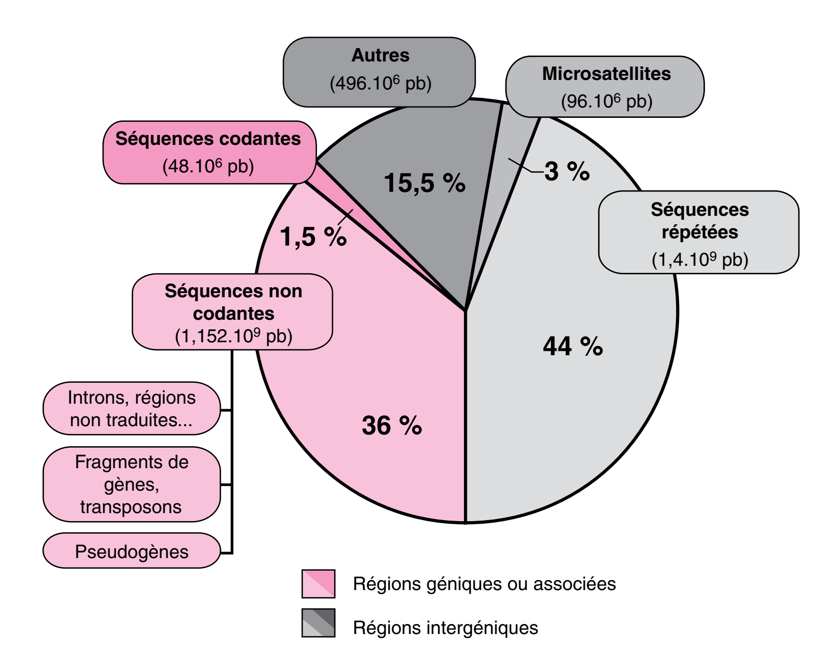


Figure 3.13 Différents types de séquences au sein du génome humain.

La majorité des séquences ne codent pas des protéines. Seules 1,5 % des séquences soit 48 mégabases sur 3 200 pour le génome complet, correspondent à des exons au sein de gènes codants.

Chaque personne humaine possède une séquence ADN unique, c'est-à-dire qui lui est propre. La séquence publiée par le consortium international correspond à la séquence combinée d'un petit nombre de donneurs anonymes et n'est donc pas exactement la séquence exploitable pour chacun d'entre nous. L'un des projets d'avenir consistera à développer la puissance des techniques de séquençage afin de pouvoir établir pour chaque personne et à faible coût l'exacte séquence de son ADN. On estime aujourd'hui qu'en moyenne l'ADN de deux humains pris au hasard différerait de 10^6 à 10^7 paires de bases dont une grande partie se présente sous la forme de **SNP** (Single Nucleotide Polymorphism). Ce polymorphisme moléculaire constitue l'une des bases « chimiques » des différences phénotypiques de tous ordres qui sont observables entre les personnes (voir chapitre 5.6).



Vers la séquence pour tous

Sera-t-il possible d'atteindre cet objectif dans un délai de temps raisonnable et à quel coût ? Le séquençage du génome humain a coûté près de 3 milliards de dollars et a mobilisé, durant près de 11 ans, plus de 1 000 chercheurs du monde entier. Aux technologies de type « Sanger » (décrites au chapitre 6 du *Mini Manuel de Biologie Moléculaire*) se sont substituées des technologies de plus en plus sophistiquées et de plus en plus miniaturisées de sorte qu'aujourd'hui il devient possible de réaliser le reséquençage d'un génome complet en à peine quelques jours, bientôt en quelques heures. Le principe de ces technologies à haut débit consiste à amplifier de courtes séquences de quelques dizaines de bases dans des volumes réactionnels extrêmement réduits afin de générer des millions de très courtes séquences d'ADN amplifiées qui sont ensuite séquencées individuellement. Ces courtes séquences sont ensuite réordonnées grâce à des outils bio-informatiques de plus en plus performants qui les comparent à une séquence de référence.

Le coût de ce reséquençage évolue également de manière vertigineuse : dès octobre 2006, la fondation américaine *X Prize Foundation* a annoncé un prix de 10 millions de dollars pour la première entreprise capable de séquencer un génome complet pour moins de 1 000 $! Cet objectif sera probablement atteint à court terme et certaines sociétés de biotechnologies annoncent déjà avoir trouvé des solutions pour réduire les coûts reséquençage à moins de 100 $.

Ces nouvelles technologies de reséquençage ouvrent des perspectives révolutionnaires en matière de recherche (*Tableau 3.4*). Il devient par exemple possible de connaître précisément les altérations du génome d'une seule cellule cancéreuse, d'analyser simultanément les gènes exprimés par l'ensemble des microorganismes d'un milieu donné (cette approche est appelée « métagénomique ») ou encore d'amplifier et de reséquencer spécifiquement des séquences génomiques ciblées de plusieurs mégabases.

En peu de temps, la démocratisation des technologies de séquençage à haut-débit, rendue possible par une réduction très importante des coûts de séquençage, a fourni à de nombreux champs d'investigation, de nouveaux outils d'analyse particulièrement puissants, agissant comme un accélérateur de recherches.



TABLEAU 3.4 EXEMPLES D'APPLICATIONS DES TECHNOLOGIES DE SÉQUENÇAGE À HAUT-DÉBIT.

Application	Exemple
Reséquençage complet de génome	Recherche de polymorphismes à l'échelle du génome entre 2 individus, recherche de nouvelles mutations, analyse fine de gènes.
Séquençage *de novo*	Alignement de séquences issues des technologies de séquençage sans utiliser de génome de référence. Cette approche reste encore problématique, notamment à cause du grand nombre de séquences répétées mais a été utilisée avec succès pour séquencer certains génomes bactériens. Il s'agit d'un enjeu majeur pour les prochaines années.
Reséquençage de régions ciblées	Identification de nouveaux polymorphismes ou de mutations causales.
Métagénomique	Identification d'une flore bactérienne ou virale au sein d'une niche environnementale particulière ; analyse globale des voies métaboliques et de l'expression des gènes d'une population microbienne particulière (ex : analyse du microbiome).
Séquençage du transcriptome	Analyse de l'expression des gènes à l'échelle du génome, identification de transcrits rares.
Séquençage des petits ARNs	Identification par séquençage de petits ARN de type micro ARN et ARN non codants (ARN transcrit mais non traduit en protéines).
Séquençage après immunoprécipitation de la chromatine (ChiP-Seq)	Cartographie, à l'échelle du génome, des sites d'interaction entre l'ADN et les protéines régulant l'expression des gènes.
Code-barre moléculaire	Séquençage simultanés de plusieurs individus, chaque échantillon multiplexé étant repéré par une séquence d'ADN unique constituant un code-barre moléculaire.
Séquençage de génomes anciens	Reconstitution et analyse de génomes anciens d'espèces disparues à partir d'ADN nucléaire, même dégradé (ex : génome du Mammouth, Homme de Neandertal).
Séquençage *de novo*	Alignement de séquences issues des technologies de séquençage sans utiliser de génome de référence. Cette approche reste encore problématique, notamment à cause du grand nombre de séquences répétées mais a été utilisée avec succès pour séquencer certains génomes bactériens. Il s'agit d'un enjeu majeur pour les prochaines années.



Une carte d'identité génétique

Avec la baisse du coût du décryptage d'un génome, de plus en plus de génomes entièrement séquencés seront disponibles pour les chercheurs. Les retombées attendues sont considérables, notamment dans le domaine de la santé humaine.

Il est probable qu'à l'avenir, chacun disposera de son propre génome et de l'information génétique qu'il contient. Un patient pourra ainsi connaître ses prédispositions à développer certains types de cancer, ou encore recevoir un traitement personnalisé de sa maladie en fonction de ses propres caractéristiques génétiques : c'est l'ère de la « génomique personnelle ».

La recherche de caractéristiques moléculaires permettant de déterminer si un individu est porteur d'une mutation responsable d'une maladie génétique est depuis longtemps l'objectif de nombreuses équipes de recherche. On trouvera ci-après (*Fig. 3.14* et *3.15*) deux exemples simples d'utilisation des séquences répétées de type microsatellites permettant de repérer des mutations génétiques.

Avec l'arrivée des technologies de génotypage à haut débit et bientôt grâce à la démocratisation des outils de séquençages à haut débit, certaines sociétés se sont spécialisées dans le dépistage de la susceptibilité à certaines maladies génétiques en exploitant la variabilité du génome. Ainsi, il est d'ores et déjà possible, contre quelques centaines de dollars et un prélèvement de salive, de connaître le risque (relatif) de développer des maladies comme la maladie d'Alzheimer, l'asthme, le cancer du sein, de la prostate ou de colon, la maladie de Crohn, l'intolérance au lactose, la sclérose en plaque, l'infarctus du myocarde, l'obésité, le psoriasis, la polyarthrite rhumatoïde ou encore les diabètes type 1 et type 2 !

Si dans certains cas médicaux, par exemple lorsqu'une atteinte familiale est suspectée, ce genre de service peut s'avérer particulièrement intéressant, il convient de rester prudent face aux dérives éthiques qui peuvent accompagner ce genre de services. En effet, sans compter l'effet catastrophique que pourrait avoir ce genre de révélations sans accompagnement médical et psychologique, il est important que ces informations restent personnelles et ne puissent être transmises à des tiers ou être utilisées commercialement… On imagine aisément l'intérêt de certains assureurs à disposer d'informations sur les risques cardio-vasculaires d'un client venant souscrire un crédit…



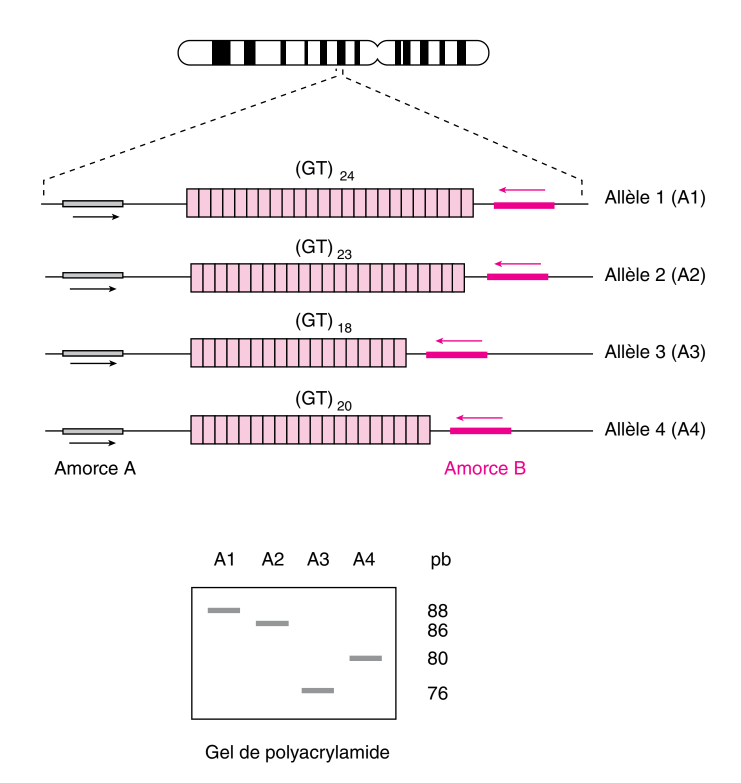


Figure 3.14 Analyse du polymorphisme d'un microsatellite.

Le microsatellite (courte séquence d'ADN répété) est ici constitué d'une répétition variable (de 20 à 24) de dinucléotides GT (CA sur le brin complémentaire). Dans les régions flanquantes du microsatellite, des amorces PCR ont été choisies pour réaliser un test d'amplification. Le schéma représente 4 variants en longueur (taille), qui sont autant d'allèles différents du microsatellite. Ils sont identifiés par leur distance de migration en gel de polyacrylamide. Pour qu'un locus microsatellite soit informatif, son polymorphisme ne doit pas être inférieur à 12 répétitions du motif nucléotidique. À un locus donné, un individu portera pour chaque paire de chromosomes homologues, l'un ou l'autre des variants du microsatellite. Dans une population, à un locus donné et pour un microsatellite, de nombreux variants sont généralement observés. Des centaines de loci polymorphes sont présents tout au long du génome de chaque individu. Par un choix raisonné de microsatellites et d'amorces PCR pour leur identification après amplification et séparation électrophorétique, des empreintes génétiques spécifiques de chaque individu de l'espèce peuvent être obtenues. C'est le principe du code-barre génétique utilisé notamment par la police scientifique.



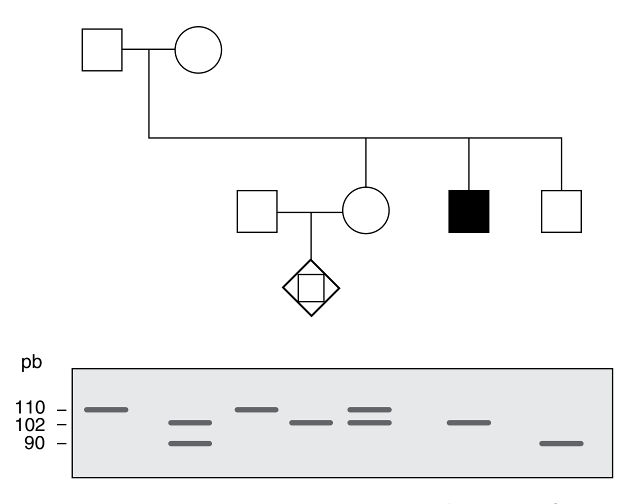


Figure 3.15 Diagnostic prénatal basé sur l'utilisation du polymorphisme d'un microsatellite.

Le microsatellite est situé dans la région promotrice du gène codant la dystrophine, une protéine importante dans la formation des muscles squelettiques. À la suite d'une réaction PCR et d'une analyse en gel de polyacrylamide, on observe chez un fœtus en cours de gestation, la présence d'un amplifiat de 102 pb révélant un haplotype identique à celui de son oncle myopathe (atteint de la maladie de Duchenne). Le risque que le sujet développe à la naissance cette maladie très invalidante est donc élevé.



POINTS CLEFS

➤ L'ADN est un acide nucléique bicaténaire, c'est-à-dire constitué de deux brins polynucléotidiques associés par des liaisons hydrogène entre les bases puriques et pyrimidiques constitutives des nucléotides. Les liaisons hydrogène s'établissent toujours entre une purine de l'un des brins et une pyrimidine de l'autre brin. Dans cet appariement complémentaire, l'adénine (A) est donc toujours associée à la thymine (T) par deux liaisons hydrogène et la guanine (G) interagit toujours avec la cytosine (C) grâce à trois liaisons hydrogènes. Ainsi, dans l'ADN, la quantité des purines (A + G) est égale à celle des pyrimidines (C + T).

➤ Le principe d'appariement complémentaire des bases a une grande valeur pratique. Il permet au généticien de déterminer la séquence nucléotidique d'un brin d'ADN à partir de celle du brin complémentaire, qu'il s'agisse



d'un brin ADN ou d'un transcrit ARN. L'orientation d'un brin polynucléotidique est définie par la présence des groupes chimiques 5′-phosphate et 3′-OH portés par les carbones 5′ et 3′ des deux désoxyriboses extrêmes. Dans la double hélice d'ADN, les deux brins complémentaires sont orientés en sens inverse. On dit qu'ils sont antiparallèles.

➤ À la surface de la molécule d'ADN, les deux brins torsadés ménagent deux sillons hélicoïdaux de largeur différente, un sillon majeur ou grand sillon et un sillon mineur ou petit sillon. Ces deux sillons, tout particulièrement le grand sillon, exposent de nombreux groupes chimiques avec lesquels interagissent dans les conditions physiologiques de la cellule des protéines dites de liaison à l'ADN. Ces protéines assument des rôles majeurs de régulation de l'expression de l'information génétique.

➤ Dans les cellules eucaryotes, à la différence des procaryotes, l'ADN est très fréquemment associé à des protéines basiques, les histones. Le résultat de l'assemblage régulier entre ADN et histones est désigné par le terme de nucléosome. La formation des nucléosomes est la première étape d'un processus qui assure l'empaquetage de l'ADN en structures compactes réduisant énormément le volume occupé par la molécule. L'ensemble est désigné sous le terme de chromatine.

➤ Certains amino-acides des queues d'histones sont parfois modifiés par des groupes acétyle et méthyle (lysine, arginine) ou phosphate (sérine). Ces modifications facilitent ou altèrent l'accessibilité de la chromatine. Elles constitueraient un « code épigénétique » qui pourrait être lu par les protéines impliquées dans la régulation de l'expression des gènes.

➤ Dans les cellules eucaryotes, le chromosome est le stade supérieur et ultime d'organisation de l'ADN. Le nombre de chromosomes et leur composition sont des caractéristiques de chaque organisme. La duplication du chromosome puis sa ségrégation s'opère dans des phases distinctes du cycle cellulaire (division cellulaire ou mitose). On distingue dans l'ordre 4 phases : G1, S, G2 et M.

➤ On désigne par le terme de génome la séquence de l'ADN correspondant à un jeu haploïde de chromosomes (*n* chromosomes). Une cellule somatique humaine, diploïde (2 *n* chromosomes), contient donc deux génomes (un génome paternel et un génome maternel) alors qu'une cellule germinale (sexuelle), haploïde (ovule ou spermatozoïde), en contient un seul. Le terme génome s'applique aussi bien à l'ADN du noyau cellulaire (génome nucléaire) qu'à l'ADN des organites (génome mitochondrial, génome chloroplastique).

➤ Le séquençage d'un génome consiste à déterminer l'ordre dans lequel sont agencés linéairement les nucléotides constitutifs de l'ADN ou de l'ARN (pour certains virus). Les plus petits génomes (appartenant à des virus) possèdent quelques milliers de paires de bases, les plus grands sont constitués de plusieurs milliards de paires de bases (certaines plantes, certains vertébrés dont les mammifères).



- La séquence établie de l'ADN humain (3, 2 milliards de pb) est aujourd'hui déposée dans les banques de données et elle est accessible pour tous par Internet. Un travail intense d'annotation destiné à comprendre comment fonctionnent les instructions contenues dans la séquence a été entrepris. Il s'annonce long et complexe car il dépend en grande partie de connaissances sur la nature exacte de ces instructions et sur leur mode d'expression.
- Dans un futur proche, la connaissance de la séquence ADN de chaque personne apportera une grande quantité de données génétiques qui seront utiles dans la prédiction des maladies, leur diagnostic mais aussi dans leur traitement.

QCM - QROC

3.1 Valider ou non les affirmations suivantes.

a) Les appariements entre les bases A, T, G, et C de l'ADN s'opèrent de façon aléatoire.

b) Il y a autant de bases T et C dans l'ADN que de bases A et G.

c) La synthèse de l'ADN s'effectue dans la cellule grâce à des ribonucléotides précurseurs.

d) L'ARN peut former des structures en double hélice.

e) Les molécules d'ARN d'une cellule sont en majorité codantes.

f) L'histone H1 est la principale protéine du nucléosome.

3.2 Les liaisons phosphodiesters des polynucléotides réunissent :

a) des bases puriques et pyrimidiques ;

b) des pentoses ;

c) des nucléosides monophosphates ;

d) des nucléosides triphosphates

3.3 L'adénine constitutive de l'ADN de deux espèces bactériennes est estimée respectivement à 30 % et 16 % de la totalité des bases. Quelles sont les proportions relatives des bases G, C, T, et A de ces deux échantillons ? L'une de ces bactéries a été isolée à proximité d'un geyser d'eau chaude proche de 70 °C. Des deux ADN analysés, lequel provient de cette bactérie ? Expliquer.

3.4 Le génome humain haploïde ($3{,}2 \times 10^9$ paires de bases environ) représente l'information contenue dans 1000 manuels de 500 pages chacun. Son support est une molécule d'ADN d'environ 1 m. Combien un manuel et une page du manuel représentent-ils de paires de bases et de longueur d'ADN ?



3.5 Sachant qu'il y a chez l'homme 23 paires de chromosomes, qu'elle est la distance moyenne en kilobases séparant deux origines de réplication ?

3.6 À combien de nucléosomes un ADN humain peut-il en théorie être associé, sachant que dans sa forme compacte un nucléosome représente un ADN de 160 paires de bases et que la distance séparant deux nucléosomes est en moyenne de 40 paires de bases ? Combien de molécules d'histone H1 et d'histone H2 seront-elles nécessaires ?

3.7 Créer une formule générale simple reliant la taille d'un ARNm, à la taille du gène correspondant, au nombre d'introns et à la taille moyenne de ces derniers.

3.8 On estime que le génome humain contient 25 000 gènes. En supposant arbitrairement que les gènes sont disposés de façon régulière et que la région codante moyenne d'un gène est de 2 kpb, quelle est la distance séparant en moyenne le centre de deux gènes ? Quelle est la proportion du génome qui code les protéines ? La nature des régions non codantes ?

RÉPONSES

3.1 **a)** non ; les appariements s'établissent obligatoirement entre les bases A et T d'une part et G et C d'autre part. **b)** oui ; **c)** non ; grâce à des désoxyribonucléotides précurseurs ; **d)** oui, de façon partielle par repliement du polynucléotide sur lui-même et à condition que la complémentarité des bases A-U et G-C soit pour l'essentiel respectée ; **e)** non, seuls les ARN messagers qui représentent 20 % environ des ARN cellulaires sont codants ; **f)** non, ce sont les histones de l'octamère H2A, H2B, H3 et H4 qui forment le nucléosome.

3.2 **c)**

3.3 L'un des ADN contient 30 % d'adénine, 30 % de thymine, 20 % de guanine et 20 % de cytosine. L'autre ADN contient 16 % d'adénine, 16 % de thymine, 34 % de guanine et 34 % de cytosine. L'ADN ayant la plus forte teneur en G + C (68 %) provient probablement de la résurgence d'eau chaude. En effet, le nombre de liaisons hydrogènes de la double hélice le rend en effet plus stable vis-à-vis de la chaleur.

3.4 Un manuel représente $3{,}2.10^6$ pb et une longueur de 1 mm ; une page du manuel représente 6 400 paires de bases et une longueur de 2 µm.



3.5 En moyenne, un chromosome contient $3{,}2 \times 10^9/23$ soit $1{,}39 \times 10^8$ paires de bases. Si l'on considère environ 3 000 origines de réplication par chromosome, la distance séparant en moyenne deux origines de réplication sera de 46 300 paires de bases.

3.6 On considère qu'à chaque ADN nucléosomal est associé un ADN de liaison, soit 160 + 40 = 200 paires de bases. Chaque ensemble sera donc représenté $3{,}2 \times 10^9/200$, soit $1{,}6 \times 10^8$ fois. Il y aura donc 160 000 000 de nucléosomes ; il y aura un nombre identique d'histones H1 et quatre fois plus d'histone H2 puisque cette dernière existe sous deux formes (H2A et H2B) et que chacune est représentée deux fois dans l'octamère d'histones.

3.7 (Taille du gène – taille de l'ARNm)/nombre d'introns = taille moyenne des introns

3.8 La distance moyenne séparant le centre de deux gènes est de 32 000 000 kpb, divisée par 25 000, soit 128 kpb. La proportion du génome humain qui code des protéines est donc de 1,56 % environ. Les autres séquences comprennent diverses régions intergéniques, répétées et non répétées, des transposons, des pseudogènes, des régions régulatrices, des introns, etc.



Introduction à la génétique des micro-organismes

PLAN

OBJECTIFS

Comprendre le transfert d'ADN chez les bactéries

Aborder les mécanismes de transduction

Comprendre l'intérêt des tests de complémentation

Intégrer les applications de la génétique des micro-organismes

Dès qu'il fut clair que des relations directes existaient entre le matériel héréditaire et les caractéristiques fonctionnelles d'une cellule ou d'un organisme, les micro-organismes s'imposèrent comme des outils de choix pour explorer en profondeur la nature des gènes. Pourtant, on ne peut pas dire *a priori* que, n'étant pas visibles à l'œil nu, les micro-organismes possèdent des caractéristiques facilement reconnaissables. Le fait d'avoir assez rapidement caractérisé leurs défauts métaboliques, isolé les souches mutantes, établi une relation entre des variants biochimiques et les mutations géniques, en un mot, engagé une véritable démarche d'étude génétique fondée sur les relations entre gènes et fonctions, a largement contribué à leur succès.

Étant donné qu'il s'agit d'organismes **haploïdes**, l'analyse génétique ne dépend pas du fait qu'une mutation soit **dominante** ou **récessive** puisqu'un seul **allèle** du gène étudié est présent. Il ne peut donc pas être masqué dans son expression par un autre allèle, comme c'est fréquemment le cas pour les organismes **diploïdes**. De plus, les micro-organismes, tout particulièrement les bactéries, peuvent produire une nouvelle génération de cellules en moyenne toutes les heures, ce qui,



pour le généticien qui étudie la transmission des caractères et des gènes au cours des générations successives, constitue un avantage considérable.

Pour toutes ces raisons et d'autres qui seront abordées au fil de ce chapitre, la génétique des micro-organismes a contribué de façon décisive à fonder les concepts majeurs de la génétique. Elle a su réaliser avant toute autre, la synthèse entre l'approche phénoménologique (les phénotypes) et l'approche moléculaire (les gènes et l'ADN).

4.1 GÉNÉTIQUE BACTÉRIENNE

Les bactéries appartiennent à une classe d'organismes appelés **procaryotes** qui inclut également les algues bleues ou **cyanobactéries**. Les procaryotes ne possèdent pas de noyau. Leurs gènes, constitués d'ADN, sont regroupés essentiellement sur un seul chromosome circulaire. Des structures de plus petite taille, les **plasmides**, portent des gènes, tels les gènes de résistance à des agents chimiques, qui fournissent à la cellule la possibilité de vivre et se multiplier dans un environnement défavorable.

Organismes modèles dont le plus connu est la bactérie *Escherichia coli* (*E. coli*), ils se divisent rapidement et se cultivent sur des milieux liquides ou solides contenant les éléments nutritifs de base (milieu minimal : sels inorganiques et une source de carbone). Cultivées sur milieu solide en boîte de Pétri, les bactéries sont immobilisées et restent regroupées. À partir de 10^7 cellules, la masse de cellules constitue une **colonie**, visible à l'œil nu. Toutes les cellules d'une colonie isolée sur une boîte sont issues d'une seule cellule, elles ont donc toutes le même matériel génétique et constituent un **clone**.

Mutants bactériens

En traitant les bactéries avec des agents mutagènes, on peut obtenir de très nombreuses mutations qui empêchent la cellule de se multiplier dans un milieu minimal. Elle ne pourra croître que si l'on ajoute à ce milieu tel ou tel métabolite dont la synthèse n'est plus assurée dans la cellule mutante. Ainsi, un grand nombre de mutations touchant la synthèse des métabolites essentiels pour la croissance bactérienne ont été identifiées, chacune d'entre elles correspondant à l'une des enzymes mises en jeu dans les étapes de biosynthèse des diverses molécules biologiques.

Pour une espèce bactérienne donnée, la souche qui a perdu la capacité de synthétiser un métabolite essentiel (un acide aminé par exemple) est dite **auxotrophe** (pour cet acide aminé). À l'inverse, la souche de type sauvage qui ne présente pas cette exigence nutritionnelle sera dite **prototrophe**. Certains mutants ne sont plus capables d'utiliser par exemple un ose particulier comme source de carbone, on parle alors



de **mutants cataboliques**. Certaines substances chimiques, comme les antibiotiques, peuvent tuer les bactéries mais certaines d'entre elles, appelés **mutants résistants**, peuvent se diviser même en présence de l'antibiotique et fonder des colonies. Tous ces divers types de mutants *(Tableau 4.1)* fournissent des **marqueurs génétiques** pour suivre l'évolution des génomes lors d'expériences.

Tableau 4.1 Symboles utilises en génétique bactérienne.

Symboles	Phénotypes associés
leu$^-$	Auxotrophe pour la leucine qui doit être en plus du milieu minimum
leu$^+$	Prototrophe pour la leucine, alors inutile dans le milieu minimum
lac$^-$	Incapable d'utiliser le lactose comme source de carbone
lac$^+$	Utilise le lactose comme source de carbone
amp^s	Sensibilité à un antibiotique, ici l'ampicilline
amp^r	Résistance à un antibiotique, l'ampicilline

Conjugaison bactérienne

En 1946, Joshua Lederberg et Edward Tatum, par une expérience à la fois simple et élégante, exploitent les exigences *d'E. coli* en certains nutriments pour démontrer l'existence de la recombinaison génétique. Ils utilisent deux souches A et B présentant des exigences nutritionnelles différentes.

La souche A est auxotrophe pour la méthionine et la biotine et prototrophe pour la leucine et la thréonine ; son phénotype est donc : met$^-$, bio$^-$, leu$^+$, thr$^+$. La souche B est auxotrophe pour la thréonine et la leucine et prototrophe pour la méthionine et la biotine ; d'où son phénotype : met$^+$, bio$^+$, leu$^-$, thr$^-$. Si des cultures de bactéries A et B en mélange sont étalées sur des boîtes contenant un milieu minimum sans supplément nutritionnel, quelques colonies apparaissent après 48 h. Seules des bactéries prototrophes (met$^+$, bio$^+$, leu$^+$, thr$^+$) sont capables de se développer sur un tel milieu. Par contre, aucune colonie n'est visible sur les boîtes témoins ensemencées avec des bactéries A ou B *(Fig. 4.1)*. Pour s'assurer que les souches ne sécrétaient pas de substances qui auraient été absorbées et utiliser par les autres cellules pour leur prolifération, Bernard Davis construisit un tube en U dont les deux bras sont séparés par un filtre qui ne peut laisser passer que les molécules dissoutes dans le milieu. En introduisant la souche A dans un bras, et la souche B dans l'autre, et après plusieurs heures

d'incubation, Davis testa les cellules de chaque bras du tube et constata l'absence de cellules de phénotype met⁺, bio⁺, leu⁺, thr⁺. Il en conclut qu'une union physique des bactéries est nécessaire pour qu'un transfert d'information génétique entre les deux souches ait eu lieu. Cette union physique, visible en microscopie électronique, est appelée **conjugaison**.

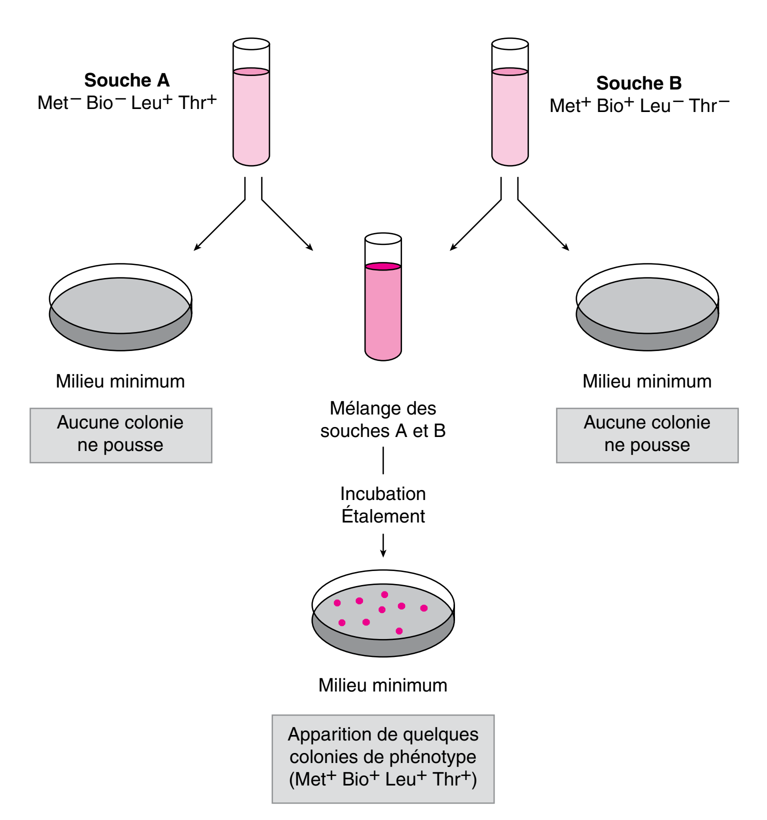


Figure 4.1 Mise en évidence du transfert de matériel génétique entre bactéries.

Les souches A et B ne peuvent pas se diviser sur un milieu minimal. En effet la souche A doit trouver dans le milieu la méthionine et la biotine et la souche B, la leucine et la thréonine. Or, le milieu minimal ne les contient pas. Le mélange des souches, qui favorise le transfert de matériel génétique, conduit à l'apparition de colonies recombinantes capables de synthétiser elles-mêmes les quatre métabolites à partir des constituants du milieu minimal.

Le facteur sexuel

L'expérience décrite ci-dessus suggère l'existence d'une sexualité chez les bactéries. En effet, il existe des cellules avec un rôle de **donneur** et des cellules ayant le rôle de **receveur**. Cependant, les génomes de



ces bactéries ne fusionnent jamais en entier pour constituer un nouveau descendant comme chez les organismes supérieurs. Le transfert de l'information génétique est unidirectionnel et est assuré par un **facteur de fertilité** ou **facteur F**. C'est un ADN circulaire autoréplicatif appelé **épisome** dont héritent les cellules filles indépendamment du chromosome bactérien *(Fig. 4.2)*. Les bactéries qui le possèdent sont dites **F⁺** assurant le rôle de donneur, et celles qui ne le possèdent pas sont dites **F⁻** et receveuse. Le facteur F comporte, entre autre, toute l'information génétique nécessaire pour son transfert d'une bactérie F⁺ vers une bactérie F⁻. Lors du transfert, un contact physique (pont cytoplasmique) s'établit entre les deux bactéries, grâce notamment à l'action de **pili** sexuels codés par le facteur F et qui se trouvent à la surface des bactéries F⁺.

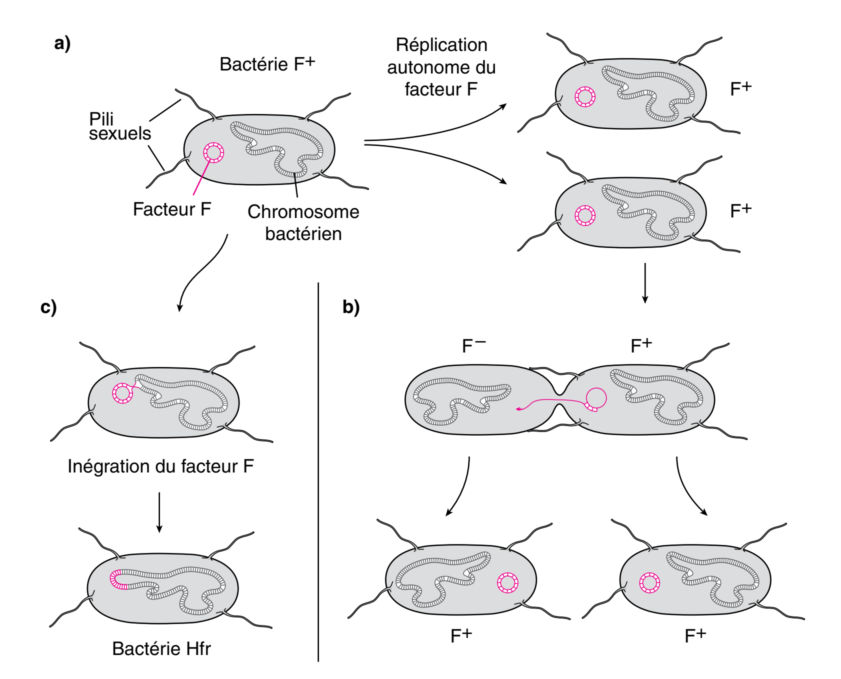


Figure 4.2 Quelques propriétés de l'épisome ou facteur F.

En **a**, présent dans les bactéries F⁺, le facteur F se réplique de façon autonome et il est hérité par toutes les cellules filles ; en **b**, lors du croisement d'une souche F⁺ avec une souche F⁻, toutes les bactéries deviennent F⁺ ; et en **c**, formation d'une bactérie Hfr.

Quand on croise des bactéries F⁺ et des bactéries F⁻ toutes les bactéries deviennent F⁺. Dans les bactéries donneuses, une copie simple-brin de l'ADN F est synthétisée selon un mécanisme particulier

appelé **réplication en cercle roulant**. Un brin de l'ADN F est transféré dans le cytoplasme de la cellule réceptrice où il va servir de matrice pour la synthèse du second brin. L'autre brin d'ADN resté dans la cellule donneuse lui permettra de la même façon de reconstituer par synthèse du brin complémentaire un facteur F fonctionnel. Dans ce type de croisement, aucun fragment du chromosome bactérien de la souche F^+ n'est transféré vers la souche F^-, on assiste seulement au transfert entier du facteur F selon des modalités respectant le principe de réplication semi-conservative *(Fig. 4.2)*. La **conjugaison bactérienne** permet le transfert de l'ADN (facteur F dans le cas présent), d'une bactérie à une autre.

Les souches Hfr

Le croisement d'une souche F^+ de type sauvage, c'est-à-dire non mutée avec une souche F^- auxotrophe pour certains métabolites et résistante à un antibiotique puis l'étalement du mélange sur le milieu de sélection (milieu minimum contenant l'antibiotique) aboutissent au développement de quelques rares colonies recombinantes témoignant du transfert de matériel génétique entre les deux souches bactériennes. Les colonies recombinantes proviennent de bactéries F^- (auxotrophes) ayant reçu par conjugaison les gènes nécessaires pour restaurer leur capacité à croître sur le milieu minimum, c'est-à-dire présentant le phénotype sauvage. Cette conjugaison exceptionnelle où quelques rares bactéries F^+ transfèrent du matériel génétique à partir de leur chromosome principal est due à l'intégration du facteur F dans le chromosome principal *(Fig. 4.2)*. Il est possible d'isoler les quelques cellules qui ont intégré le facteur F dans leur chromosome.

Si l'expérience de croisement est alors répétée avec la nouvelle population constituée de cellules ayant intégré le facteur F dans leur chromosome, on obtient 1000 fois plus de recombinants. Ces souches sont appelées pour cette raison **Hfr** pour Haute fréquence de recombinaison.

Grâce à l'intégration du facteur F dans le chromosome principal, ces souches peuvent transférer efficacement aux souches F^- une partie de leur génome. De ce fait, la bactérie réceptrice peut devenir, à cause du fragment de chromosome reçu, homologue de son propre chromosome, une cellule partiellement diploïde. Cette information génétique transférée s'intègre ensuite dans le chromosome de la bactérie F^- par un double crossing-over *(Fig. 4.3)*, ce qui permet d'obtenir des cellules et des colonies recombinantes stables. Les cellules F^- qui portent un allèle du donneur ont donc participé à la conjugaison et sont appelées **exconjugants**.

Bactérie Hfr sensible à la streptomycine Str^S

Transfert du matériel génétique a^+

Bactérie F^- résistante à la streptomycine Str^r

Intégration de a^+ par double crossing-over

Dégradation du fragment d'ADN linéaire

Bactéries F^- (a^+ Str^r)

Milieu + streptomycine

Sélection de recombinants pour le caractère a^+

Figure 4.3 Principe de l'intégration de marqueurs génétiques dans le chromosome bactérien et sélection des cellules recombinantes.

La bactérie F^- est auxotrophe pour un acide aminé (a-) et résistante à la streptomycine (Str^r). La bactérie Hfr est prototrophe (a^+) et sensible (Str^s). Le transfert orienté vers la bactérie F^- de la région chromosomique (a^+) contenant le(s) gène(s) permettant la synthèse de l'acide aminé, suivi d'un double crossing-over, entraîne la création d'une souche recombinante (a^+, Str^r) qui peut croître et former des colonies sur un milieu minimal contenant de la streptomycine, milieu qui n'autorise ni la croissance de la souche Hfr (Str^s), ni celle de la souche F^- (a^-).

Il existe un grand nombre de souches de type Hfr puisque l'intégration du facteur F s'effectue grâce à un mécanisme de crossing-over en de nombreux sites du chromosome de la bactérie *(Fig. 4.4)*. Les régions d'appariements, identiques sur le plasmide et le chromosome, proviennent d'éléments mobiles, appelés séquences d'insertion. Toutes ces souches Hfr transfèrent donc à des souches F^- toujours la partie de leur chromosome qui est immédiatement en amont de l'origine de transfert. En effet, l'origine et la direction du transfert sont déterminées par la même extrémité du facteur F, appelée **origine de transfert** libérée par clivage. Lors de la conjugaison d'une bactérie Hfr avec une

bactérie F⁻, l'information génétique du facteur F sera donc toujours transférée en dernier lieu après que tout le chromosome de la souche Hfr soit transféré, c'est-à-dire très rarement. Le processus est fragile et la rupture du pont cytoplasmique et de l'ADN en cours de transfert se produit fréquemment. Par conséquent, la souche F^- n'est quasiment jamais convertie en F^+ ou en Hfr à la suite de son croisement avec la souche Hfr.

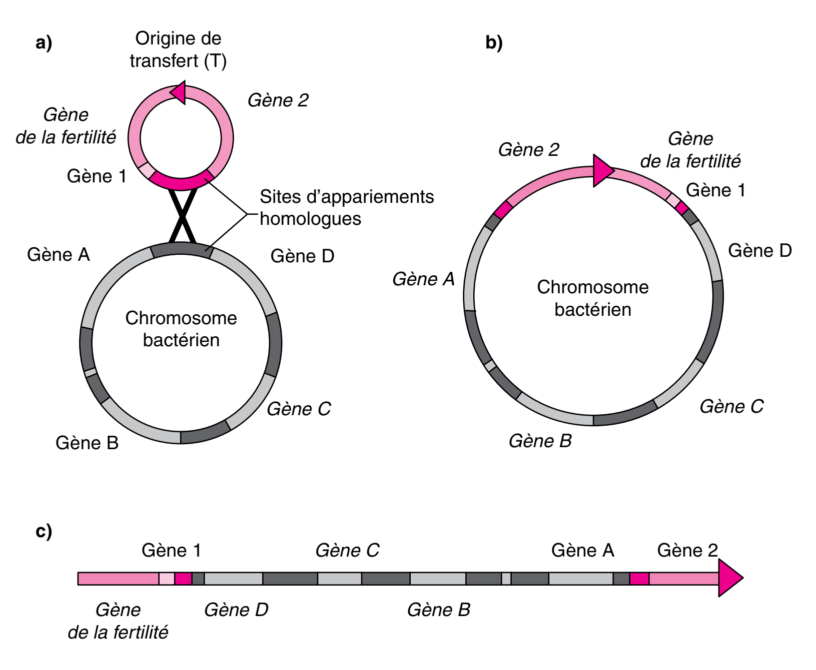


Figure 4.4 Intégration du facteur F dans le chromosome bactérien et transfert orienté de marqueurs génétiques.

En **a**, le facteur F s'intègre dans le chromosome bactérien par crossing-over en mettant en jeu des homologies de séquences entre les deux molécules d'ADN. En **b**, le facteur F intégré dans le chromosome d'une bactérie devenue de ce fait Hfr ; en **c**, dans le cas présent l'endroit où s'intègre le facteur F permet le transfert à partir de l'origine T des marqueurs dans l'ordre, A, B, C et D (1 et 2 sont des marqueurs symboliques pour orienter l'ADN inséré).

Bien que le facteur F s'intègre de façon stable dans le chromosome de la souche Hfr, parfois il quitte le chromosome par inversion exacte du processus de recombinaison qui lui a permis son insertion. Parfois, sa sortie n'est pas rigoureusement identique à l'insertion, le plasmide emportant avec lui une partie du chromosome bactérien. Le nouveau plasmide qui en résulte s'appelle **plasmide F'**. Les facteurs F' peuvent



porter plusieurs gènes chromosomiques et sont nommés d'après les gènes supplémentaires qu'ils portent.

Cartographie par conjugaison

La *figure 4.5* décrit le principe d'une expérience de conjugaison interrompue chez les bactéries. Au bout de 5 min de conjugaison entre la souche Hfr et la souche F^-, aucun recombinant ne sera obtenu pour les marqueurs génétiques testés quel que soit le crible de sélection utilisé. Après 15 min de conjugaison, on peut caractériser des recombinants de la souche F^- pour a^+, b^+. Le crible de sélection des recombinants

Figure 4.5 En **a** : Principe d'une expérience de conjugaison interrompue chez les bactéries. On a représenté le passage du chromosome de la bactérie Hfr dans la bactérie F^- et la rupture du pont cytoplasmique et du chromosome en fonction du temps. L'arrêt de la conjugaison s'effectue par agitation. Une partie du fragment d'ADN ayant pénétré dans la bactérie receveuse peut s'intégrer par recombinaison homologue dans le chromosome. En **b**, chromosomes de deux souches Hfr. L'origine de transfert est représentée par un triangle, le terminus de transfert par un rectangle noir. La position et l'orientation de l'origine de transfert imposent le sens du transfert. Pour la souche 1 c'est le gène *pur* qui sera transféré en premier alors que pour la souche 2 c'est le gène *his*.



est très important : si au bout de 15 min de conjugaison, on sélectionne les recombinants uniquement pour le caractère a^+ parmi tous les recombinants a^+ obtenus, certains seront a^+b^+ mais on n'aura jamais de recombinant pour b^+ uniquement. Autrement dit, le premier crossing-over dans ces conditions a eu lieu obligatoirement du côté de l'origine de transfert et le deuxième soit entre a^+ et b^+ (recombinant a^+) soit en dehors de l'intervalle séparant a^+ de b^+ (recombinant a^+b^+). Il ne faut pas oublier que deux crossing-over sont nécessaires pour l'échange d'information génétique entre un fragment linéaire d'ADN (fragment transféré) et un ADN circulaire (chromosome bactérien).

Cette expérience type peut être exploitée pour déterminer la liaison entre les gènes. Pour qu'un exconjugant intègre dans son chromosome les gènes du donneur de manière définitive, le fragment chromosomique du donneur doit recombiner avec le chromosome du receveur. Il faut noter que les unités sont des minutes et non des fréquences de recombinaison. Il s'avère néanmoins possible d'utiliser la fréquence de recombinaison pour la cartographie du chromosome bactérien. La recombinaison se produit ici entre un génome complet issu de F^-, appelé l'**endogénote**, et celui incomplet, appelé l'**exogénote**, issu du donneur Hfr. À ce stade la cellule possède deux copies d'un fragment d'ADN, c'est un diploïde partiel ou **mérozygote**. Un seul crossing-over dans un mérozygote ouvrirait le cercle et il n'y aurait pas de recombinant viable. Pour que le chromosome demeure circulaire, il faut obligatoirement un nombre pair de crossing-over.

Pour déterminer la liaison et obtenir des distances cartographiques significatives, il est nécessaire d'établir des conditions dans lesquelles tous les marqueurs ont une probabilité égale d'être transmis, de sorte que les fréquences de recombinaison ne dépendent que de la distance entre les gènes considérés. De ce fait, pour étudier la liaison génétique entre les marqueurs a^+, b^+, c^+, d^+, et e^+ par exemple, il faut sélectionner les recombinants pour le dernier marqueur qui sera transféré, c'est-à-dire e^+ dans le cas présent *(Fig. 4.5)*. Pourquoi ? Parce que nous savons que chaque cellule qui a reçu un fragment de chromosome portant le dernier marqueur a également reçu les marqueurs antérieurs (a^+, b^+, c^+ et d^+). Nous pouvons alors calculer la distance entre marqueurs qui est égale au pourcentage de crossing-over dans les intervalles qui les séparent. Pratiquement, cela revient à calculer parmi le nombre total de recombinants obtenus pour un marqueur déterminé le pourcentage de ceux qui l'ont été grâce à un crossing-over entre ce marqueur et le marqueur voisin.

En combinant l'utilisation de diverses souches Hfr portant les mêmes marqueurs génétiques (locus d'auxotrophie pour tel ou tel acide aminé



par exemple), on constate que l'ordre de transfert des marqueurs n'est pas conservé d'une souche à l'autre, ce qui s'explique par la circularité du chromosome bactérien d'une part, l'origine et la direction du transfert d'autre part. En répétant l'expérience avec une autre souche Hfr qui transfère le marqueur b^+ en premier, celui qui sera transféré le dernier ne sera plus e^+ mais a^+. De plus, en jouant sur le temps de transfert, qui peut être contrôlé en provoquant la rupture du pont cytoplasmique et de l'ADN par une vigoureuse agitation, on a pu cartographier, c'est-à-dire déterminer la position relative des loci d'auxotrophie, la résistance à des antibiotiques, à l'infection phagique, en un mot, cartographier le chromosome bactérien.

Les plasmides R

La résistance multiple aux antibiotiques est certainement la propriété la plus alarmante des bactéries pathogènes, mise en évidence dans les années 1950, avec des bactéries du genre Shigella responsables de la dysenterie. La bactérie a très rapidement acquis une résistance à de nombreux antibiotiques (pénicilline, tétracycline, streptomycine, etc.) couramment utilisés pour enrayer la maladie. Ce phénotype de résistance multiple transmis sous la forme d'un groupe génétique unique, peut se transmettre non seulement à d'autres souches de Shigella mais également à des espèces apparentées. Cette propriété, proche de la mobilité du facteur F chez *E. coli*, est très favorable pour la bactérie pathogène car la résistance se propage très rapidement à travers une population. Les vecteurs transportant les multiples résistances d'une cellule à une autre appartiennent à un groupe de plasmides, les **plasmides R** qui sont transférés rapidement lors de conjugaisons cellulaires. Les premiers découverts, les plasmides R de Shigella, font partie d'une grande série de plasmides de même type. Tous présents à l'état plasmidique dans le cytoplasme des bactéries, ils portent de nombreux types différents de gènes comme les gènes de résistance aux métaux lourds (plasmide R6), de production de bactériocines (plasmide Col E1), de tumorigénicité dans les plantes (plasmide T1 chez *Agrobacterium tumefaciens*).

La transformation bactérienne

Le mécanisme par lequel une bactérie absorbe de l'ADN exogène à partir de son environnement et l'intègre de façon fonctionnelle dans son chromosome est appelé **transformation**. Ce mécanisme de modification des propriétés génétiques par de l'ADN en solution fut découvert chez *Streptococcus pneumoniae* en 1928 par Frederick Griffith et Oswald Avery. Par la suite, en 1944, Avery, Mac Leod et Mac Carty démontrèrent que le principe transformant était bien de l'ADN.



Compétence génétique naturelle chez les bactéries

Tout d'abord observée chez la bactérie Gram-positive, *S. pneumoniae*, la transformation se révéla un processus naturel chez de nombreuses bactéries Gram-positives et Gram-négatives. La capacité qu'ont les bactéries à absorber naturellement des fragments d'ADN exogène en solution, processus nommé **compétence naturelle**, est génétiquement définie. En effet, les bactéries **compétentes**, expriment des protéines spécialisées qui permettent aux cellules de capter puis d'intégrer des fragments d'ADN présents occasionnellement dans leur environnement *(Fig. 4.6)*.

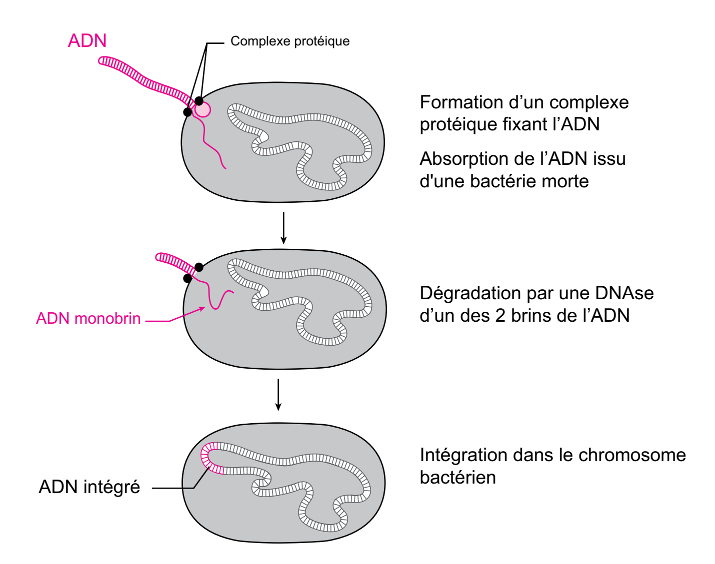


Figure 4.6 Principe de la transformation bactérienne. Un fragment d'ADN libéré par une bactérie morte est absorbé. Au niveau des protéines du complexe de fixation de l'ADN, et au cours de son entrée un des brins de l'ADN est dégradé, l'autre brin envahit l'ADN de la cellule receveuse. Celui-ci sera par la suite, après synthèse du brin complémentaire, intégré dans le chromosome bactérien.

La fixation de l'ADN chez *Neisseria gonorrheae* et *Haemophilus influenzae* est sélective dans la mesure où le fragment incorporé doit contenir une séquence nucléotidique spécifique. Chez *H. influenzae,* il doit contenir une séquence consensuelle de 29 pb avec un segment hautement conservé de 9 pb. Aujourd'hui, grâce au séquençage du génome de cet organisme, on sait que cette séquence est fréquemment représentée. Au total, 1 465 séquences de ce type ont été identifiées dans le génome de *H. influenzae* qui a une taille de 1 830 kpb. C'est une sorte de signature, d'identité, du génome de cet organisme. Pour



Streptococcus pneumoniae et *Bacillus subtilis*, l'absorption de l'ADN n'est pas sélective. Aucune séquence spécifique n'est requise pour transformer ces deux bactéries. Cependant, chez les deux, le développement de la compétence est régulé par une sorte de communication entre les cellules : avant de développer son état de compétence, une cellule doit recevoir un signal des autres cellules de la population. Un tel système permettrait aux cellules de ne développer un état de compétence qu'au voisinage d'autres cellules de la même espèce et par conséquent, de n'absorber que de l'ADN propre à cette espèce.

La transformation artificielle

Les mécanismes de transformation naturelle décrits ci-dessus mettent en jeu un grand nombre de protéines, et parfois une séquence d'ADN spécifique. Au vu de la complexité de ce système de transformation naturelle, l'application de la technologie de l'ADN recombinant à ces micro-organismes reste relativement difficile. D'autres bactéries comme *E. coli*, sont plus facilement transformables par de l'ADN exogène. L'état de compétence de ces bactéries peut être accru en les traitant par des ions divalents comme le Ca^{2+}. On parle dans ce cas de **transformation artificielle**. La transformation artificielle induite par Ca^{2+} n'est efficace chez *E. coli* que lorsque l'ADN est ajouté sous une forme capable d'autoréplication. Les ADN de plasmides et les chromosomes viraux intacts ont une grande efficacité de transformation, alors que les fragments linéaires d'ADN bactérien transforment très faiblement les bactéries.

Ainsi, la transformation bactérienne qui révéla au monde que l'ADN est le support du matériel génétique, nous fournit aujourd'hui un moyen puissant pour l'étude des gènes *in vivo* et *in vitro*. Le mécanisme de transformation des bactéries par des plasmides permet théoriquement d'introduire n'importe quel fragment d'ADN dans une cellule pour étudier les conséquences de son expression.

Cartographie du chromosome bactérien grâce à la transformation

Nous avons vu que la transformation des bactéries par l'ADN en solution permettait de transférer un caractère d'une bactérie à une autre.

Considérons deux bactéries, l'une réceptrice, l'autre donatrice, différentes par plusieurs caractères : soit *a*, *b*, *c*, *d*, les gènes gouvernant les caractères distinctifs de la bactérie réceptrice et *A*, *B*, *C*, *D*, ceux de la bactérie donatrice. Pour chacun des caractères, on observe que la fréquence d'apparition des bactéries transformées portant un seul caractère nouveau est de l'ordre de 10^{-2}.

Les fréquences des bactéries transformées simultanément pour deux caractères (double transformation) sont en général égales au produit



des fréquences de chacune des transformations simples (par exemple pour *CD* $10^{-2} \times 10^{-2} = 10^{-4}$). En revanche pour *BC* cette fréquence est plus élevée $0{,}2 \times 10^{-2}$. L'explication est que les gènes *B* et *C* sont suffisamment proches sur le chromosome bactérien pour être portés par un même fragment de telle sorte que les doubles transformés *BC* résultent de la pénétration dans la bactérie réceptrice d'un unique fragment (gènes liés). En appliquant cette technique, on a pu établir la carte du chromosome de *B. subtilis*.

4.2 GÉNÉTIQUE DES BACTÉRIOPHAGES

Le cycle biologique des bactériophages

C'est par nécessité biologique que les virus sont des parasites. Ils ne possèdent pas la capacité de synthétiser leurs composants de façon indépendante. Pour se reproduire, ils doivent introduire leur génome dans une cellule hôte vivante. Tous les virus sont de petites particules formées d'une enveloppe protectrice riche en protéines qui facilite leur transport d'une cellule à l'autre. Leur matériel génétique est soit de l'ADN, soit de l'ARN mais jamais les deux en même temps.

Les **bactériophages** (ou phages) sont des virus qui infectent les bactéries. Ils ont joué, et continuent de jouer, un rôle fondamental dans la génétique bactérienne et la génétique moléculaire en général. Pendant plus de trente ans (1940-1973), ils ont été des outils précieux pour étudier les conséquences de l'introduction d'un nouveau matériel génétique dans une cellule. Ainsi, les études menées sur les phages ont aidé à la compréhension des structures et de l'expression des gènes, l'assemblage de structures biologiques complexes et le transfert de matériel génétique entre bactéries.

Les phages les plus étudiés sont ceux qui se multiplient chez *E. coli* ou des bactéries très proches. On les a arbitrairement nommés T1, T2, T4, P1, F1, M13 ou λ par exemple. Un phage infecte une bactérie en s'attachant à un récepteur spécifique de la surface bactérienne. Une fois fixé, le phage injecte son matériel génétique dans le cytoplasme. Certains des gènes du phage sont exprimés immédiatement (gènes précoces) après l'injection de l'ADN dans la bactérie, en utilisant des enzymes présentes dans le cytoplasme de la bactérie. Ces gènes précoces codent des enzymes nécessaires à la réplication de l'ADN phagique en un grand nombre de copies. Les gènes tardifs commencent ensuite à s'exprimer. Ce sont principalement des gènes qui codent les protéines nécessaires à la production de nouvelles capsides phagiques. Une fois les différents constituants phagiques assemblés, la paroi bactérienne est détruite (on parle de **lyse** bactérienne), libérant un grand nombre de nouveaux phages dans le milieu environnant. Chacun d'entre



eux est capable d'infecter une nouvelle bactérie, ce qui permet un nouveau cycle biologique du phage *(Fig. 4.7)*.

Libération des bactériophages
ADN phagique
Bactériophage λ libre
Infection d'une bactérie
a)
Cycle lytique
Lyse de la cellule
ADN λ
Intégration dans le génome bactérien
b)
Cycle lysogénique
Multiplication bactérienne
Prophage intégré
ADN λ
Induction
Libération du génome phagique
ADN λ
Génome bactérien
ADN λ
(Prophage)
Répresseur du bactériophage λ
ADN λ
(Prophage)
Prolifération des bactéries

Figure 4.7 Cycle lytique d'un bactériophage **(a)**, et cycle lysogénique **(b)**.

La lysogénie

Dans les années 1920, on a observé que certaines souches bactériennes apparaissaient résistantes à certains phages. Cependant, ces bactéries résistantes entraînaient la lyse d'autres bactéries lorsqu'on mélangeait



les cultures. Ces bactéries résistantes sont dites **lysogènes**. Vers le milieu des années 1940, André Lwoff montra qu'une bactérie lysogène est une bactérie qui contient un génome viral, inséré en un point déterminé de son propre chromosome. La présence de l'ADN du phage à l'état intégré dans le chromosome bactérien, on parle de **prophage**, protège la bactérie de l'infection par un autre phage (surinfection). Cette immunité reste stable et peut être transmise aux cellules bactériennes filles. Le prophage se duplique par conséquent au même titre que les autres régions du chromosome bactérien. Chez une petite fraction de cellules lysogènes, on peut assister à la production de phages infectieux. On parle de l'**induction** du prophage des bactéries lysogènes. Ce processus prive, évidemment, la cellule de son immunité, lyse la bactérie et libère les phages infectieux dans le milieu. Ces derniers peuvent infecter toute bactérie non lysogène présente dans le milieu de culture.

Des études ultérieures montrèrent que différents agents, tels que le rayonnement ultraviolet ou certains produits chimiques, pouvaient induire la lyse d'une fraction importante d'une population de bactéries lysogènes.

Ainsi, on a pu classer les phages en deux catégories :

- les phages **virulents** : ils provoquent toujours la lyse des bactéries et ne peuvent exister à l'état de prophage ;
- les phages **tempérés** : ils suivent fréquemment un cycle lysogénique et s'installent dans la bactérie hôte sous forme de prophage. Dans certaines conditions, ils peuvent suivre un cycle lytique, lorsque le prophage est induit.

La transduction

En 1951, J. Lederberg et N. Zinder, travaillant sur la bactérie *Salmonella*, montrèrent que des caractères pouvaient être transmis de façon passive d'une bactérie à une autre par l'intermédiaire d'un bactériophage. Ils appelèrent ce phénomène la **transduction**.

Une expérience typique de transduction consiste à infecter une bactérie leu^+ par un phage transducteur spécifique de cette bactérie. Le développement de ce phage dans les bactéries entraîne leur lyse, ce qui libère un grand nombre de descendants. La majorité de ces phages est identique au phage infectant. Exceptionnellement, la capside renferme un fragment d'ADN du chromosome bactérien intégré dans la molécule d'ADN du phage. Quand on infecte une culture de bactéries leu^- par le lysat contenant l'ensemble des phages, certains de ceux qui ont « embarqué » un petit fragment de chromosome bactérien introduisent dans les bactéries réceptrices le gène *leu*$^+$. Par recombinaison, l'allèle *leu*$^+$ peut se substituer à l'allèle leu^- et la bactérie se trouve transduite pour ce caractère. Tout fragment de l'ADN de la bactérie



donneuse portant un marqueur génétique pourra se trouver transduit de la sorte. On parle alors de transduction **généralisée** *(Fig. 4.8)*.

Il existe un autre type de phages transducteurs qui ne transportent qu'une partie restreinte et bien déterminée du chromosome bactérien du fait d'une insertion du transducteur en une position précise du

LeuB
Infection d'une bactérie Leu+
Phages sauvages (très nombreux)
LeuB
Phage ayant intégré un morceau d'ADN bactérien contenant le gène *LEUB* (rare)
LeuB
Leub
Infection d'une bactérie Leu−
LeuB
Leub
Intégration du fragment dans le chromosome
LeuB
Bactérie transduite portant le caractère Leu+

Figure 4.8 Mécanisme de la transduction généralisée. Le gène *leuB* code la β-isoproylmalate déshydrogénase, enzyme impliquée dans la voie de biosynthèse de la leucine. L'allèle muté de ce gène est noté *leub*.



chromosome bactérien. Ce phénomène est appelé transduction **spécialisée**. Le phage λ, phage tempéré, constitue un bon exemple de transducteur spécialisé. L'induction du prophage λ de la bactérie *E. coli* lysogène entraîne la production de certaines particules défectueuses due à une excision incorrecte du prophage. Le génome du phage emporte avec lui un fragment d'ADN du chromosome d'*E. coli* alors que certains gènes phagiques restent dans le chromosome bactérien *(Fig. 4.9)*. Cet ADN phagique anormal, portant des gènes bactériens *gal* ou *bio* proches du site d'intégration (on appelle ces phages λdgal pour λ-défectueux gal, ou λdbio), peut être empaqueté et infecter d'autres bactéries. Si une particule phagique normale infecte également la bactérie (double infection) λdgal peut s'intégrer dans le génome de l'hôte au site spécifique de fixation du phage λ. Ainsi, les gènes gal sont transduits chez la deuxième bactérie infectée.

Les séquences appelées *att* (« attachement ») contenant les sites spécifiques sur l'ADN phagique (*attP*) et sur le chromosome bactérien (*attB*) ne contiennent qu'une très courte région d'homologie. La réaction d'intégration est catalysée par l'enzyme **Int** (intégrase) synthétisée par le bactériophage λ de façon abondante au début de l'infection. Participe également à cette catalyse une protéine accessoire d'*E. coli* appelée **IHF** (Integration Host Factor : facteur d'intégration à l'hôte).

L'intégration du phage λ est réversible et son excision nécessite une seconde enzyme **Xis** (excisionnase) synthétisée elle aussi par le phage. La formation d'un complexe entre l'intégrase et **Xis** catalyse l'excision du prophage, ce qui permet de reconstituer un ADN bactérien et un ADN phagique intacts *(Fig. 4.9)*.

La transduction généralisée peut être exploitée pour cartographier le chromosome bactérien, c'est la **cartographie par transduction**. Supposons que l'on souhaite déterminer les positions relatives des marqueurs génétiques a^+, b^+, c^+ sur le chromosome bactérien. On peut faire croître par exemple le phage transducteur P1 en infectant une souche a^+, b^+, c^+, puis avec les phages obtenus, on infecte une souche a^-, b^-, c^-. Si l'examen des phages transduits pour le marqueur a^+ montre que 50 % d'entre eux porte également b^+ et 2 % c^+, on pourra interpréter ce résultat comme signifiant que le marqueur a^+ est plus proche de b^+ que de c^+.

Si maintenant on sélectionne les transduits pour le marqueur c^+ et qu'on observe, parmi eux, 3 % contenant a^+ et aucun b^+, cela indique que c^+ est plus proche de a^+ que de b^+. Comme la première expérience montre que b^+ est plus proche de a^+ que ne l'est c^+, le marqueur a^+ se trouve sur le chromosome bactérien entre b^+ et c^+, mais plus proche de b^+ que de c^+. Le principe du raisonnement est basé sur le fait que plus deux marqueurs sont proches l'un de l'autre, plus ils ont



ADN λ
attP
Crossing-over
Chromosome bactérien
gal +
attB
bio +
gal +
attB
prophage λ
attB
bio +
Formation d'une boucle
Excision correcte
Formation d'une boucle
Excision incorrecte
gal+
bio+
gal+
bio+
ADN λ
attP
λ dgal
gal+
+
+
gal+
attB
bio+
bio+

Figure 4.9 Intégration du phage λ et mécanisme de la transduction spécialisée.

Le bactériophage λ s'intègre toujours dans la même région entre *gal* et *bio* par crossing-over. Les séquences *attP* contenant les sites spécifiques sur l'ADN phagique et attB sur le chromosome bactérien contiennent une très courte région d'homologie qui permet la recombinaison. L'induction du prophage entraîne l'excision normale de l'ADN du phage λ. Rarement, le phage λ peut s'exciser incorrectement et emporter avec lui un fragment d'ADN du chromosome bactérien (gal^{+} dans le cas présent). Les particules phagiques ainsi formées sont déficientes car certains gènes λ sont restés dans le chromosome de l'hôte, on les appelle : λ*dgal*. De la même façon, on parle de λ*dbio*. Cependant, il est tout à fait possible d'utiliser ces particules pour transduire une autre bactérie en réalisant une cotransduction avec un phage λ sauvage.

de chances d'être cotransduits. L'étude de plusieurs marqueurs génétiques permet par cette manière qui utilise la fréquence de recombinaison, de cartographier une région du chromosome bactérien, c'est-à-dire d'ordonner les marqueurs les uns par rapport aux autres.

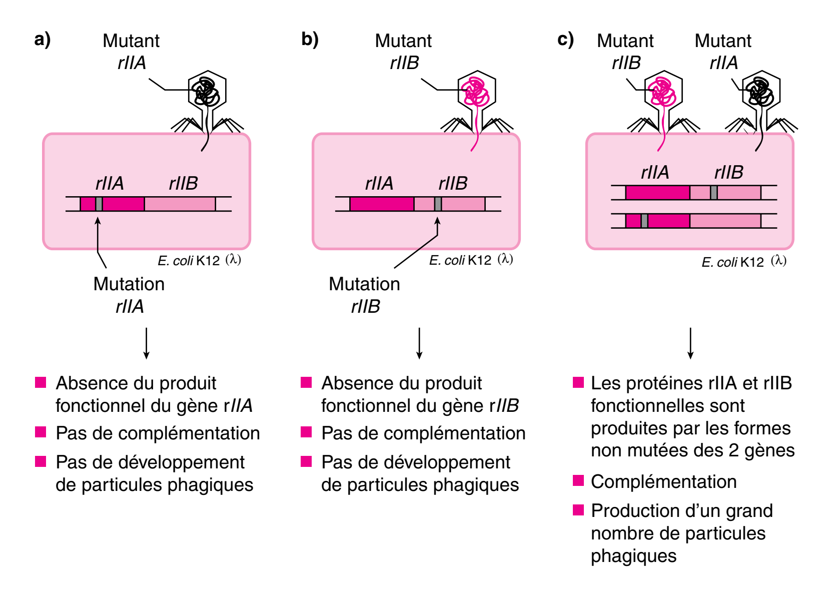


4.3 TEST DE COMPLÉMENTATION OU TEST D'ALLÉLISME FONCTIONNEL

Avec la découverte que les gènes étaient constitués d'une succession de nucléotides organisés en double hélice, il devint très clair qu'une telle entité pouvait être définie par un ensemble de sites mutationnels détectés au même locus et engendrant le même phénotype mutant. Cependant, les approches expérimentales décrites ci-dessus ne permettaient pas d'établir les limites d'un gène. Le test qui s'est révélé être un bon indicateur des limites d'un gène était celui de la complémentation.

Les travaux génétiques réalisés sur le locus *rII* du phage T4, par Benzer (1961), constituent un bon exemple non seulement pour comprendre ce test de complémentation, mais aussi pour aborder la structure fine du gène. Le locus *rII* du phage T4 est formé de deux gènes contigus (*rIIA* et *rIIB*) dont les produits régulent la durée du cycle du phage. Les plages de lyse produites par le phage sauvage T4, lorsqu'on l'inocule sur la souche d'*E. coli B,* se distinguent facilement de celles engendrées par l'infection de la même souche B, soit par des mutants *rIIA* soit par des mutants *rIIB*. Dans le premier cas (phage sauvage), on observe la formation de petites plages de lyse alors que celles-ci sont nettement plus grandes dans le second cas (phages mutants). Il existe une autre souche, *E. coli K12 (λ)* sur laquelle seul le phage T4 de type sauvage peut se multiplier en produisant d'ailleurs le même phénotype de plages que sur la souche B. En revanche, les mutants *rIIA* ou *rIIB* arrivent à injecter leur ADN dans *E. coli K12 (λ)* mais ne peuvent pas se reproduire *(Fig. 4.10)*. Par contre, lorsque *E. coli K12 (λ)* est infectée simultanément par deux mutants *rIIA* et *rIIB* (**croisement entre phages**), on s'aperçoit alors que l'infection est suivie de l'apparition, sur un milieu nutritif gélosé, d'un grand nombre de plages de lyse. Ceci démontre que les gènes *rIIA* et *rIIB* assurent des fonctions différentes (deux gènes différents), chacune nécessaire à la multiplication sur *E. coli K12 (λ).* De façon plus générale, cette expérience indique que deux génomes réunis dans une même cellule peuvent se complémenter l'un l'autre lorsqu'ils sont mutés dans des gènes différents (le verbe complémenter est un néologisme devenu nécessaire en génétique). Ceci constitue un test de complémentation. Par exemple, lors de l'infection par le seul mutant *rIIB*, le produit du gène *rIIA* est synthétisé et fonctionnel mais l'absence du produit fonctionnel du gène *rIIB* empêche le développement du phage. De la même manière, quand on infecte par le seul mutant *rIIA,* la fonction du produit du gène *rIIB* est assurée mais elle est insuffisante pour le développement du phage en l'absence de la fonction du gène *rIIA (Fig. 4.10)*.





- Absence du produit fonctionnel du gène r*IIA*
- Pas de complémentation
- Pas de développement de particules phagiques

- Absence du produit fonctionnel du gène r*IIB*
- Pas de complémentation
- Pas de développement de particules phagiques

- Les protéines rIIA et rIIB fonctionnelles sont produites par les formes non mutées des 2 gènes
- Complémentation
- Production d'un grand nombre de particules phagiques

Figure 4.10 Représentation schématique du test de complémentation pour l'étude du locus *rII* chez le phage T4.

a) Infection par un mutant *rIIA* ; **b)** Infection par un mutant *rIIB* ; **c)** Infection mixte par les deux mutants *rIIA* et *rIIB*, il y alors complémentation et production de particules phagiques, les deux gènes sauvages fonctionnels *rIIA* et *rIIB* étant présents dans la même bactérie. Dans tous les cas, l'ADN bactérien n'est pas représenté.

Jusqu'aux années 1940, le test de complémentation a été largement pratiqué, notamment pour savoir si deux mutants ayant le même phénotype et obtenus indépendamment étaient affectés dans la même fonction (même gène) ou dans des fonctions différentes (deux gènes différents) sans les connaître obligatoirement. S'il y avait une complémentation entre les deux mutations réunies dans une cellule (phénotype sauvage), on concluait que les deux mutations affectaient deux gènes différents. Au contraire, s'il n'y avait pas de complémentation (phénotype mutant), on concluait que le phénotype observé, obtenu indépendamment, était dû à la mutation d'un même gène. Ainsi l'absence de complémentation révélée par ce test génétique est une preuve d'allélisme qui indique que les mutations sont localisées dans les limites d'un gène mais cette preuve ne renseigne pas sur la localisation des mutations les unes par rapport aux autres à l'intérieur d'un gène : elles peuvent occuper exactement le même emplacement ou, fréquemment, des emplacements différents.



4.4 TEST D'ALLÉLISME STRUCTURAL

Lorsqu'on s'est aperçu que la recombinaison, longtemps considérée comme n'ayant lieu qu'entre deux gènes, pouvait avoir lieu à l'intérieur même d'un gène, l'ancien test de complémentation fut remplacé par deux tests. Le test d'allélisme fonctionnel ou test de complémentation indique si deux mutations appartiennent ou non au même gène, c'est-à-dire si elles sont ou non alléliques. Le **test d'allélisme structural** révèle que des mutations appartenant au même gène, donc alléliques, ne complémentent pas mais peuvent cependant subir une recombinaison.

La preuve sans équivoque que la recombinaison peut avoir lieu à l'intérieur même d'un gène a été fournie par l'étude des gènes *rIIA* et

Figure 4.11a La sélection de recombinants intragéniques du gène *rIIA* du phage T4.

Deux mutants du même gène *rIIA* (*rIIA1* et *rIIA2*) sont utilisés après leur développement séparé dans *E. coli* B pour infecter la souche *E. coli* K12 (λ). Ces infections simples ne révèlent aucun développement phagique sur *E. coli* K12 (λ).



b

Mutant *rIIA1*
Mutant *rIIA2*
rIIA1
rIIA2
E. coli B

Pas de recombinaison
Recombinaison (événement rare)

rIIA1
rIIA2
E. coli B

Crossing-over
rIIA1
rIIA2
E. coli B

rIIA1
rIIA2
E. coli B

rIIA1 *rIIA2*
rIIA
Type sauvage
E. coli B

Lyse des bactéries
Libération de phages

rIIA1
E. coli K12 (λ)
Absence de plage de lyse

rIIA2
E. coli K12 (λ)
Absence de plage de lyse

rIIA1 *rIIA2*
E. coli K12 (λ)
Absence de plage de lyse

rIIA
E. coli K12 (λ)
Présence de plage de lyse

Figure 4.11b Sélection de recombinants intragéniques du gène *rIIA* du phage T4.

Deux mutants du même gène *rIIA* (*rIIA1* et *rIIA2*) sont mis ensemble en contact de la bactérie après leur développement simultané dans *E. coli* B (obtention de recombinants sauvages potentiels). Les phages obtenus sont utilisés pour infecter *E. coli K12* (λ). On constate l'apparition de quelques plages de lyse qui témoignent d'une recombinaison intragénique entre les deux allèles mutés reconstituant le gène *rIIA* sauvage. L'infection par les simples mutants *rIIA1* ou *rIIB* tout comme par les doubles mutants ne révèlent aucun développement phagique sur *E. coli K12* (λ)



rIIB du bactériophage T4 en utilisant le test d'allélisme structural *(Fig. 4.11)*. La cotransfection de la souche d'*E. coli B* par deux mutants *rIIA1* et *rIIA2* permet d'obtenir un lysat de particules phagiques. La mise en évidence d'une recombinaison **intragénique** (à l'intérieur du même gène) lors de cette cotransfection est basée sur l'infection de la bactérie *E. coli K12 (λ)* avec les phages produits par la souche B. En effet, seuls des recombinants ayant acquis le type sauvage peuvent se développer sur *E. coli K12 (λ)*. Dans cette expérience, on obtient quelques plages de lyse (phages sauvages) pour un grand nombre de croisements testés, au contraire du résultat obtenu lors du test de complémentation. Cette expérience a permis de montrer l'apparition de phages de type sauvage en croisant deux phages mutés au niveau d'un même gène ce qui confirme que la recombinaison est tout à fait possible à l'intérieur d'un même gène. On parle de **recombinaison intragénique** *(Fig. 4.11)*. L'étude d'un grand nombre de mutants des gènes *rIIA* et *rIIB* révèle un large éventail de fréquences de recombinaison intragénique. Cette gamme de valeurs indique que certaines mutations peuvent être plus ou moins rapprochées, ce qui permet de construire une carte génétique précise et de définir une structure fine pour chacun des gènes.

4.5 GÉNÉTIQUE DE LA LEVURE

Cet eucaryote unicellulaire présente un cycle de reproduction **haplo-diplobiontique** *(Fig. 4.12)*, tout à fait intéressant pour les généticiens.

Ainsi, de très nombreux mutants ont été obtenus avec la levure *Saccharomyces cerevisiae* dont les étapes du cycle de reproduction sont totalement contrôlées au laboratoire par les généticiens. Des mutants affectés dans le cycle cellulaire ont pu être isolés chez la levure. Bien évidemment, les gènes intervenant dans le cycle cellulaire sont des gènes essentiels ou indispensables, il a donc été nécessaire de caractériser des mutants **conditionnels** qui présentent un phénotype sauvage viable dans des conditions dites **permissives** et un phénotype muté létal dans des conditions dites **non permissives**. Les généticiens ont eu l'idée de rechercher des mutants **thermosensibles** (*ts*). À température permissive (20 °C), les levures présentent un phénotype sauvage, elles se divisent. À température non permissive (37 °C), elles présentent un phénotype mutant, elles ne se divisent pas.

Les mutations qui augmentent ou diminuent la fonction du produit d'un gène sont extrêmement utiles pour l'étude d'un système cellulaire. L'effet de la mutation (phénotype) donne des indications sur la protéine elle-même et sur la voie dans laquelle elle agit.

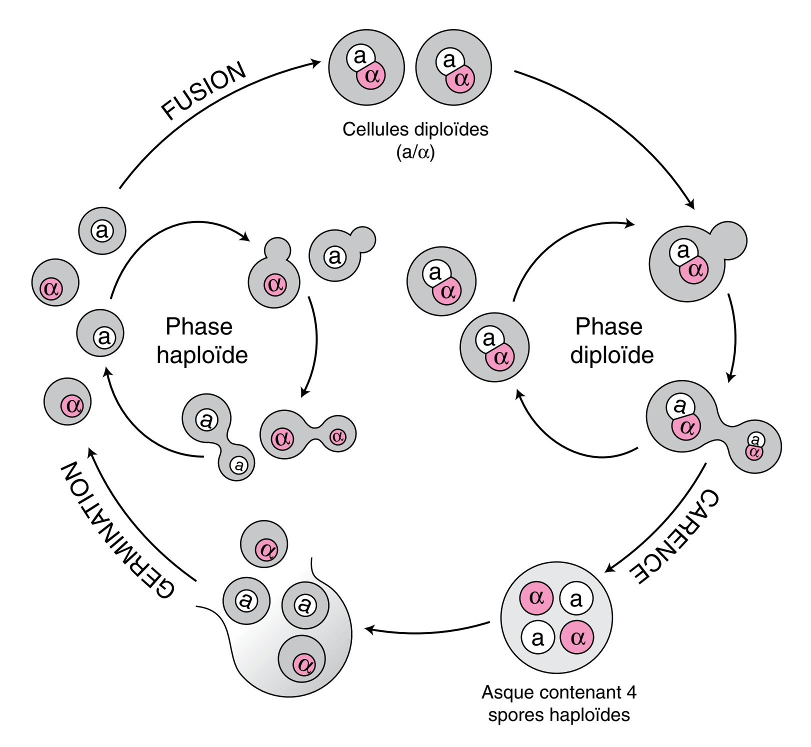


Figure 4.12 Cycle de reproduction haplodiplobiontique de la levure *Saccharomyces cerevisiae.*

La levure peut proliférer par reproduction végétative, en se divisant par bourgeonnement, aussi bien à l'état haploïde qu'à l'état diploïde. Le passage à l'état diploïde s'effectue par fusion de deux cellules haploïdes de signes sexuels opposés (*a* et α). Une cellule diploïde peut subir la méiose et donner 4 spores, qui après germination donneront 2 colonies haploïdes de signe sexuel a et 2 colonies haploïdes de signe sexuel α.

Les mutants peuvent être obtenus par :

- mutagenèse au hasard (agents mutagènes), elle a pour but de lier une fonction à un ou plusieurs gènes. En général, on cible le génome dans son entier ;
- mutagenèse *in vitro* sur un gène donné afin d'identifier les domaines fonctionnels d'une protéine ;
- mutagenèse ciblée, un gène est spécifiquement inactivé ou altéré (voir p. 116).

Après avoir identifié les différents mutants, leur caractérisation s'effectue par :

- l'étude phénotypique détaillée avec d'éventuels cribles secondaires ;



- la détermination du caractère dominant ou récessif de la mutation ;
- la vérification du nombre de mutations liées au phénotype (ségrégation 2/2 de la mutation après croisement avec la souche parentale) ;
- on catégorise les mutants par groupes de complémentation.

Les groupes de complémentation et dénombrement des gènes

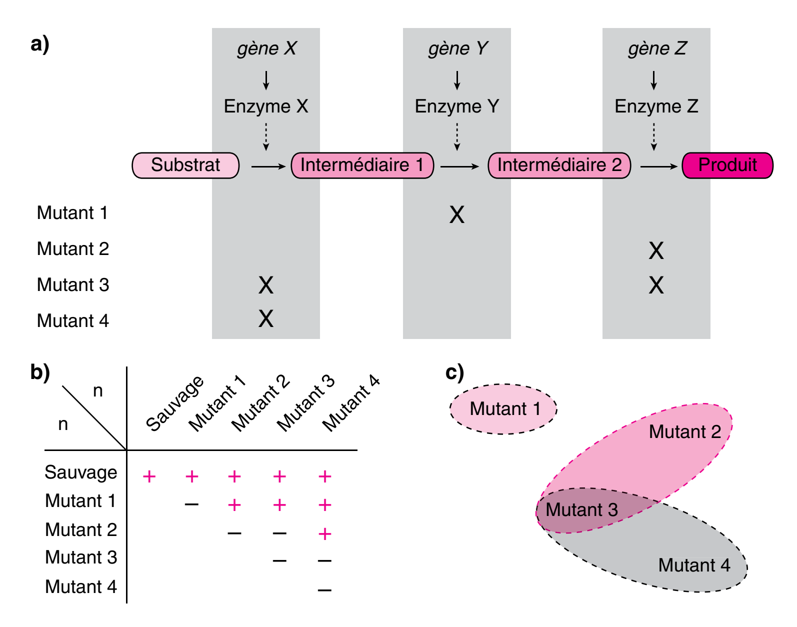
La génétique de la levure est essentiellement basée sur la possibilité de croiser deux souches haploïdes portant des mutations différentes pour former une souche diploïde. L'induction de la sporulation du diploïde permettra l'obtention d'une **tétrade** dont les spores isolées par dissection avec un micromanipulateur formeront chacune une colonie. Par le passé, ces croisements ont permis d'établir les cartes chromosomiques en utilisant la fréquence de recombinaison entre deux mutations comme mesure de la distance génétique. Les croisements sont utilisés aujourd'hui pour générer des souches avec de nouvelles combinaisons de mutations et pour analyser les relations génétiques entre différentes mutations.

L'analyse génétique d'une voie métabolique ou physiologique consiste à dénombrer le nombre de gènes impliqués dans celle-ci, à identifier la fonction biochimique pour chacun d'eux, voire leurs interactions éventuelles, etc. En obtenant le maximum de mutants pour une voie étudiée, on s'assure de « toucher » tous les gènes impliqués dans cette voie, à l'exception des gènes dont la mutation serait létale. En croisant entre eux tous les mutants récessifs, on construit des groupes de complémentation. Un groupe de complémentation est constitué d'un ensemble de mutants ne complémentant pas entre eux et qui de ce fait ont le même gène muté. Ainsi, si :

- les deux mutations sont dans le même gène, la souche diploïde aura le même phénotype mutant que les haploïdes ;
- les deux mutations sont dans deux gènes différents, le diploïde aura un phénotype sauvage, chaque haploïde apportant une copie de l'autre gène sauvage.

Le nombre de groupes de complémentation obtenu par l'analyse fonctionnelle des croisements entre les divers mutants de la voie étudiée indique le nombre minimal de gènes impliqués dans cette voie *(Fig. 4.13)*.





n \ n	Sauvage	Mutant 1	Mutant 2	Mutant 3	Mutant 4
Sauvage	+	+	+	+	+
Mutant 1		–	+	+	+
Mutant 2			–	–	+
Mutant 3				–	–
Mutant 4					–

Figure 4.13 Détermination du nombre de groupes de complémentation et par analogie du nombre minimal de gènes impliqués dans une voie métabolique **(a)**. Les différents mutants sont croisés entre eux et avec un haploïde sauvage. Les diploïdes obtenus sont mis en culture, les résultats obtenus sont reportés dans le tableau **(b)**. On observe trois groupes de complémentation **(c)**, on peut en conclure qu'au moins trois gènes différents sont impliqués dans cette voie de biosynthèse. (–) pas de complémentation, donc probablement les mutations sont dans le même gène pour les deux mutants ; (+) complémentation, donc les mutations affectent deux gènes distincts.

Invalidation de gènes chez la levure

L'**invalidation de gènes** est basée sur la désorganisation sélective d'un gène. Chez la levure, cette technique est facilitée, par la relative fréquence de la recombinaison homologue. De plus, son cycle de reproduction haplodiplobiontique offre la possibilité de travailler sur des souches haploïdes, ce qui ne nécessite pas l'obtention d'une invalidation homozygote. Chez la levure, cette technique permet entre autres de classer les nouveaux gènes en deux catégories. Par l'analyse génétique d'une souche diploïde où l'invalidation du gène étudié est à l'état hétérozygote *(Fig. 4.14)*, on peut déterminer si le gène en question est essentiel à la viabilité ou non.

De la même manière on peut introduire un gène X au locus du gène *URA3* qui sera alors invalidé. L'invalidation du gène *URA3* qui code l'oritidine5'-phosphate décarboxylase, enzyme impliquée dans la synthèse de l'uracile (Ura3), permet de contre-sélectionner le marqueur



a)

Plasmide

Gène X

URA3

Séquences du *gène X*

URA3

Fragment d'ADN linéarisé

Transformation d'une souche de levure diploïde

Invalidation *in vitro* du *gène X* par insertion du marqueur de sélection *URA3*

URA3

Gène

Souche de levure hétérozygote

URA3

Gène X

Gène X

Invalidation d'un des deux allèles par recombinaison homologue

b)

Gène essentiel

Gène non essentiel

2 spores sur 4 forment une colonie de levures

Germination

Sporulation (méiose)

Sporulation (méiose)

Germination

4 spores sur 4 forment une colonie de levures

Souche de levure hétérozygote pour l'invalidation

Figure 4.14 Invalidation de gène chez la levure.

En **a**, dans une première série d'expérimentations, le gène est invalidé en le remplaçant par un marqueur de sélection (ici, *URA3*). Puis l'intégration au génome de la levure est effectuée par une recombinaison homologue, basée sur un double crossing-over. En **b**, l'analyse génétique de la descendance de la souche de levure hétérozygote pour l'invalidation montre, suivant le cas, si le gène invalidé est essentiel ou non. S'il est essentiel, les 2 spores haploïdes contenant le gène invalidé ne forment pas de colonies.

URA3 en présence d'acide 5-fluoro-orotique (FOA) qui est converti par Ura3 en 5-fluorouracil, toxique pour la cellule. Les cellules $URA3^+$ sont donc tuées en présence de 5-FOA tandis que les cellules ayant intégré le gène X sont de phénotype ura3 et ont acquis la résistance à la FOA.



Les levures eucaryotes modèles et outils

Les levures sont considérées comme des organismes modèles. On anticipe qu'une partie au moins des systèmes cellulaires fonctionne de façon similaire chez la levure et chez l'homme et par extension chez tous les eucaryotes. Ceci est vrai et de nombreux processus mis en évidence ou étudiés chez la levure se sont révélés en partie conservés chez les mammifères bien qu'avec des niveaux de complexité plus élevés comme :

- la régulation du cycle cellulaire ;
- la transmission des signaux extracellulaires ;
- les régulations de l'expression des gènes ;
- le trafic intracellulaire ;
- le vieillissement…

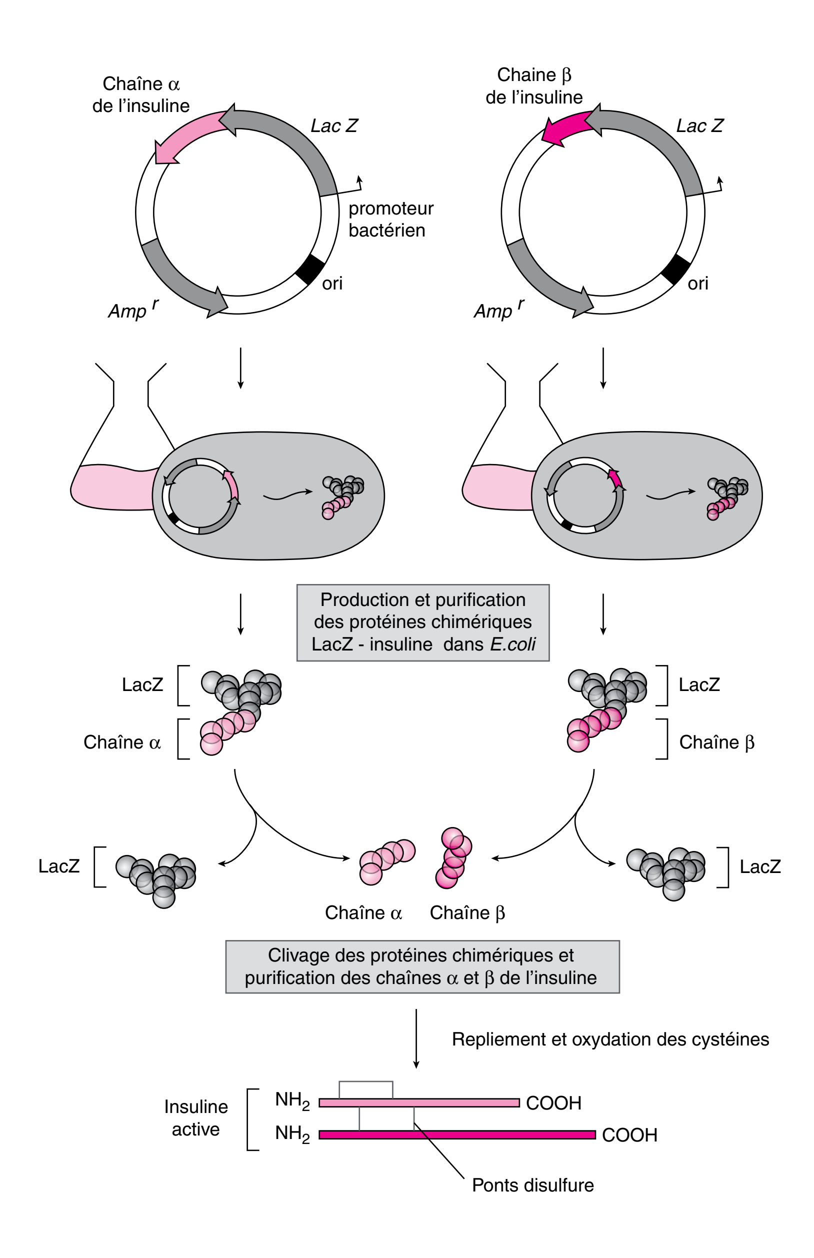
Pour le vieillissement, on a pu établir un parallèle entre la durée de vie prédéterminée de la cellule mère qui meurt après 40 à 50 générations et celle de la plupart des cellules d'eucaryotes supérieurs et identifiés des gènes impliqués dans le vieillissement. Par exemple, des mutations du gène, *WRN* (Werner's syndrom) chez l'homme et de son homologue *SGS1* chez la levure provoquent des sénescences précoces. La collection de mutants de levure a été mise à profit pour le clonage de gènes hétérologues, et ainsi de nombreux gènes de mammifères ou de plantes ont été clonés par complémentation de mutants de levure (voir dans la même collection l'ouvrage *Biologie moléculaire*).

4.6 MICRO-ORGANISMES ET GÉNIE GÉNÉTIQUE

La technologie de l'ADN recombinant utilisant différentes cellules hôtes a permis la biosynthèse de grandes quantités de protéines d'intérêt comme des protéines humaines qui n'étaient jusque là disponibles qu'en

Figure 4.15 Production d'insuline recombinante humaine par *E. coli.* ➤

L'insuline est composée de deux sous-unités α et β. Chacune d'elles est produite séparément. Dans un vecteur d'expression, portant un marqueur de sélection, le gène de résistance à l'ampicilline (*Ampr*) on introduit la séquence codant l'une ou l'autre des chaînes en aval du gène *Lac Z* placé sous la dépendance de promoteur bactérien. Les bactéries sont transformées et mises en culture sur un milieu contenant de l'ampicilline. Les clones résistants à l'ampicilline sont utilisés pour une culture en masse. La protéine chimérique *Lac-Z* sera synthétisée dans les bactéries. Après arrêt de la culture les bactéries sont lysées et les protéines purifiées. Le clivage des protéines chimériques s'effectue par traitement au bromure de cyanogène qui rompt la liaison peptidique entre la méthionine de *Lac Z* et la phénylalanine porté par la chaîne d'insuline. Les chaînes α et β sont purifiées, se replient et des ponts disulfure s'établissent entre les cystéines portées par les deux sous-unités. La molécule d'insuline est active et peut être injectée pour traiter des diabètes de type I.


Chaîne α
de l'insuline
Lac Z
promoteur
bactérien
ori
Amp r
Chaine β
de l'insuline
Lac Z
ori
Amp r
Production et purification
des protéines chimériques
LacZ - insuline dans E.coli
LacZ
Chaîne α
LacZ
Chaîne β
LacZ
Chaîne α
Chaîne β
LacZ
Clivage des protéines chimériques et
purification des chaînes α et β de l'insuline
Repliement et oxydation des cystéines
Insuline
active
NH2
COOH
NH2
COOH
Ponts disulfure



très petites quantités. La première protéine qui fût exprimée est la somatostatine, un peptide neurotransmetteur de 14 acides aminés. Depuis, de nombreuses **protéines recombinantes** ont vu le jour avec comme application première le traitement de maladies immunologiques, neurologiques, infectieuses, génétiques, etc. Un des succès de la technique est la fabrication de vaccins contre les maladies infectieuses. Elle consiste à produire la protéine de surface du pathogène qui servira alors d'antigène. Le premier vaccin réussi est celui de l'hépatite B, la protéine immunogène étant produite chez la levure. Des molécules recombinantes sont même retrouvées dans des compositions de détergents et de produits alimentaires.

La stratégie générale de production de ces différentes molécules consiste à cloner le gène d'intérêt dans un vecteur, appelé **vecteur d'expression**, qui est ensuite transféré dans une cellule qui peut assurer son expression. Les vecteurs dérivent des plasmides présents naturellement dans les bactéries et les levures. Les systèmes d'expression les plus utilisés sont procaryotes, *E. coli* et *B. subtilis*, et eucaryotes, levures, cellules d'insectes et cellules de mammifères. Les procaryotes ne sont pas souvent de bons candidats pour la production de protéines humaines notamment quand celles ci nécessitent des modifications post-traductionnelles pour être actives.

Généralement, les molécules recombinantes sont produites soit pour augmenter la quantité d'une molécule déjà existante et extraite de tissus animaux. L'insuline par exemple était purifiée à partir de pancréas de porcs sacrifiés à l'abattoir. La première insuline recombinante humaine est apparue sur le marché pharmaceutique en 1982. Cette insuline a pu être produite à l'échelle industrielle, et à des coûts moindres grâce à l'utilisation de bactéries et des techniques de l'ADN recombinant *(Fig. 4.15)*. Elle permet de traiter les patients atteints de diabètes de type I. Depuis un grand nombre de « protéines médicaments » sont produites par des bactéries comme des hormones de croissance, l'interféron, des interleukines, etc.

Le génie génétique permet d'obtenir aussi des protéines dont la structure naturelle est modifiée afin d'améliorer ou de modifier leur fonction. À titre d'exemple, la subtilisine, protéase à sérine utilisée dans les détergents, est sensible aux autres produits employés dans le blanchiment à cause de l'oxydation de la méthionine 222. Par mutagenèse dirigée, plusieurs formes mutées de la subtilisine ont été produites où la méthionine a été remplacée par d'autres acides aminés. Il a été ainsi possible de sélectionner un mutant (possédant l'alanine 222) moins actif que le sauvage mais insensible à l'oxydation.





POINTS CLEFS

- Les procaryotes ne possèdent pas de noyau. Leurs gènes sont regroupés essentiellement sur un seul chromosome circulaire.
- Les plasmides, structures de petite taille, portent des gènes, tels les gènes de résistance à des agents chimiques, qui fournissent à la bactérie la possibilité de vivre et se multiplier dans un environnement défavorable.
- Toutes les bactéries d'une colonie isolée sur une boîte sont issues d'une seule cellule, elles ont donc toutes le même matériel génétique et constituent un clone.
- Lors de la conjugaison bactérienne, le transfert d'ADN est unidirectionnel de la bactérie donneuse vers la bactérie receveuse.
- Au cours de la conjugaison, un contact physique s'établit entre les deux bactéries grâce notamment à l'action de pili sexuels codés par le facteur F.
- Le facteur F est un plasmide qui peut s'intégrer dans le chromosome bactérien, la bactérie est alors nommée Hfr pour Haute fréquence de recombinaison.
- L'origine et la direction du transfert sont déterminées par la même extrémité du facteur F, appelée origine de transfert. Lors de la conjugaison d'une bactérie Hfr avec une bactérie F^-, l'information génétique du facteur F sera toujours transférée en dernier lieu après transfert de tout le chromosome de la souche Hfr.
- Par des expériences de conjugaison bactérienne interrompue, on a pu faire la cartographie du chromosome bactérien.
- Les vecteurs transportant les multiples résistances d'une cellule à une autre appartiennent à un groupe de plasmides, les plasmides R.
- Le mécanisme par lequel une bactérie absorbe de l'ADN exogène à partir de son environnement et l'intègre de façon fonctionnelle dans son chromosome est appelé transformation.
- La capacité qu'ont les bactéries à absorber naturellement des fragments d'ADN exogène en solution, processus nommé compétence naturelle, est génétiquement définie.
- Les bactériophages (ou phages) sont des virus qui infectent les bactéries.
- Un prophage est constitué par l'ADN d'un phage à l'état intégré dans le chromosome de l'hôte bactérien.
- De courtes séquences d'ADN bactérien peuvent être transmise de façon passive d'une bactérie à une autre par l'intermédiaire d'un bactériophage. C'est la transduction.

- Certains phages transducteurs ne transportent qu'une partie restreinte et bien déterminée du chromosome bactérien du fait d'une insertion du transducteur en une position précise du chromosome bactérien. Ce phénomène est appelé transduction spécialisée.
- Le test d'allélisme fonctionnel ou test de complémentation indique si deux mutations appartiennent ou non au même gène, c'est-à-dire si elles sont ou non alléliques.
- Les mutants conditionnels présentent un phénotype sauvage viable dans des conditions dites permissives et un phénotype muté létal dans des conditions dites non permissives.
- Un groupe de complémentation est constitué d'un ensemble de mutants ne complémentant pas entre eux et qui de ce fait ont le même gène muté.
- L'invalidation de gènes est basée sur la désorganisation sélective d'un gène.
- Les levures sont considérées comme des organismes modèles. Une partie au moins des systèmes cellulaires fonctionne de façon similaire chez la levure et chez tous les eucaryotes.
- La production de protéines recombinantes s'effectue par insertion de la séquence codant la protéine d'intérêt dans un vecteur d'expression qui est ensuite transféré dans une cellule compétente qui peut assurer son expression.

QCM - QROC

6.1 Quelle est la bonne affirmation ?

a) Les gènes bactériens sont portés par un seul chromosome linéaire.

b) Les bactéries *gal*⁻ utilisent le galactose comme source de carbone.

c) Les gènes de résistances aux antibiotiques sont utilisés comme marqueurs.

d) L'auxotrophie correspond à la capacité d'une bactérie à proliférer sur milieu minimum.

6.2 Quelles propositions sont exactes :

a) Le facteur F n'est pas intégré dans le chromosome d'une bactérie Hfr.

b) Les pili sexuels relient deux bactéries F^+ lors de la conjugaison.

c) Le transfert d'ADN d'une bactérie Hfr vers une bactérie F^- est unidirectionnel.

d) C'est sous la forme monobrin que l'ADN est transféré lors de la conjugaison.



e) L'ADN transféré à la bactérie F^- s'intègre de façon aléatoire dans son chromosome.

6.3 En quoi la transduction diffère de la transformation ?

6.4 Le phage dans le cas de la transduction spécialisée :
a) emporte des régions spécifiques du génome bactérien ;
b) encapside la totalité de son ADN ;
c) porte le nom du gène bactérien qu'il a emporté ;
d) ne peut intégrer son ADN au chromosome de l'hôte que s'il y a une double infection nécessitant la présence d'une particule phagique normale.

6.5 Peut-il y avoir recombinaison entre les génomes des virus bactériens ?

6.6 Quelle est la bonne proposition :
a) Un groupe de complémentation est défini comme un ensemble de mutants qui n'ont pas les mêmes gènes mutés.
b) Le nombre de groupes de complémentation obtenu après croisement des mutants entre eux permet de définir le nombre maximum de gènes présents dans une voie métabolique étudiée.
c) Chez la levure le croisement de mutants ayant chacun une mutation récessive donnera un phénotype sauvage au diploïde hétérozygote.
d) Si les deux mutations sont dans le même gène, le diploïde aura le même phénotype mutant que les haploïdes.

6.7 Chez *E. coli*, 3 souches Hfr, dérivées de la même souche F^+, transfèrent une série de marqueurs dans l'ordre suivant :
– Hfr1 : ACBDF ;
– Hfr2 : LKJFD ;
– Hfr3 : SEMLK.
Donner l'ordre des marqueurs sur le chromosome circulaire de la bactérie F^+ dont dérivent les 3 souches Hfr.

6.8 Un bactériophage possède les gènes dans l'ordre suivant : *g1*, *g2*, *g3*, *g4*, *g5*, *g6*, *g7* et *g8* ; alors que le son prophage les a dans un autre ordre *g6*, *g7*, *g8*, *g1*, *g2*, *g3*, *g4* et *g5*. Quelle information vous apporte cette permutation ?

RÉPONSES

6.1 c) Vraie ; faux : **a)** le chromosome est circulaire, **b)** la bactérie ne métabolise pas le galactose, **d)** l'auxotrophie traduit l'incapacité à synthétiser des métabolites qui doivent être ajoutés au milieu de culture.



6.2 **c)** et **d)** vraies ; faux **a)** le facteur F est intégré au chromosome de la bactérie Hfr, **b)** ils relient les bactéries Hfr ou F^+ à une bactérie F^-, **e)** l'ADN s'intègre par recombinaison homologue.

6.3 Les phages servent d'intermédiaire pour le transfert d'ADN d'une bactérie à une autre ; la transformation concerne des fragments d'ADN libre dans le milieu.

6.4 **a)**, **c)** et **d)** vraies, **b)** faux, l'intermédiaire en boucle qui se forme entraîne une excision erronée, certains gènes phagiques restent dans le chromosome bactérien.

6.5 Oui, si deux types de phages légèrement différents infectent simultanément une bactérie hôte (voir *figure 4.11b*).

6.6 **c)** Vraie ; faux **a)** un groupe de complémentation est défini comme un ensemble de mutants qui ont le même gène muté, **b)** de définir le nombre minimum de gènes présents dans une voie métabolique, **d)** si les deux mutations sont dans le même gène, mais suffisamment distantes, il peut y avoir recombinaison homologue qui aboutira à un allèle sauvage et un allèle portant les deux mutations.

6.7 Le 1er marqueur à être transmis est le plus proche de l'origine de transfert dont l'orientation impose également le sens du transfert. Ainsi, la souche Hfr1 transfère en premier le marqueur *A* puis les autres marqueurs ; la souche Hfr2 transfère elle en premier le marqueur *L*. On constate que *F* et *D* sont communs aux deux souches, mais dans un ordre inverse. Donc, pour la souche Hfr2, l'origine de transfert est inversée par rapport à Hfr1. On observe également une inversion des marqueurs communs *L* et *K* transmis par Hfr 2 et 3. En orientant les séquences de ces 2 souches par rapport à Hfr1, l'ordre des marqueurs sur la souche d'origine est : *ACBDFJKLMES*.

6.8 On en déduit la position du site d'attachement *Att* qui permet l'intégration de l'ADN phagique par recombinaison homologue. Il se situe entre les gènes *g5* et *g6*.





Expression des gènes et des génomes

PLAN

OBJECTIFS

- Étudier les différents aspects de l'expression des gènes
- Accorder une attention spéciale aux modifications épigénétiques
- Aborder certains aspects liés aux analyses à grande échelle
- Aborder les aspects intégrés de la génétique et des autres disciplines de la biologie

On désigne par le terme « expression des gènes » le processus moléculaire et cellulaire qui consiste à convertir la séquence ADN des gènes en protéines fonctionnelles. Les différentes étapes de ce processus sont modulables depuis l'étape de la transcription de l'ADN en ARN jusqu'aux modifications post-traductionnelles des protéines. On désigne habituellement ces modulations par le concept de « régulation des gènes ». Il est à la base de la compréhension du fonctionnement courant des cellules et organismes. Les mécanismes mis en jeu dans cette régulation sont également la cible de changements mutationnels dont les effets se manifestent à la fois dans le contrôle du métabolisme et dans l'évolution des êtres vivants.

5.1 EXPRESSION DES GÈNES : LA TRANSCRIPTION DE L'ADN

Les protéines nécessaires à la transcription

Les mécanismes de la **transcription** reposent sur l'activité d'une famille d'enzymes, les ARN polymérases, qui réalisent le même type de réaction (une synthèse d'ARN à partir d'une matrice d'ADN) quelle que soit



la cellule procaryote ou eucaryote considérée. Les ARN polymérases se fixent à des régions de l'ADN désignées sous le terme de **promoteurs**, souvent avec l'aide de protéines additionnelles. Elles catalysent ensuite par différentes étapes bien définies la polymérisation des ribonucléotides triphosphates précurseurs en un polynucléotide monocaténaire dont la séquence est complémentaire du brin d'ADN ayant servi de matrice modèle (**brin transcrit**). L'autre brin d'ADN qui n'est pas transcrit est dit parfois **brin codant** car il possède une séquence identique à celle du transcrit d'ARN. Les ARN polymérases adoptent une structure tridimentionnelle en forme de pince ce qui leur permet d'enserrer le brin d'ADN (orienté dans le sens 3' → 5') et en coulissant de le transcrire en un ARN orienté dans le sens 5' → 3'. Bien que le mécanisme général chez les eucaryotes s'apparente à celui des procaryotes, il s'en distingue cependant par une intervention beaucoup plus complexe de nombreux facteurs de transcription indispensables à la formation spécifique de l'association entre ARN polymérases et promoteurs et au démarrage efficace de la transcription. On distingue à cet effet **les facteurs généraux de la transcription,** suffisants pour reconstituer le processus *in vitro*, et les **facteurs additionnels** organisés en **complexes médiateurs** indispensables à la transcription *in vivo*.

Les ARN polymérases s'attachent aux promoteurs de façon faiblement spontanée et la transcription peu active, **constitutive**, qui en résulte se situe à un **niveau basal**. Un répresseur occupant un **opérateur**, c'est-à-dire un site chevauchant le promoteur sera alors suffisant pour inhiber la transcription. À l'inverse, un activateur occupant son site souvent à proximité du promoteur aidera la fixation de la polymérase *(Fig. 5.1)*. Ce recrutement de l'enzyme par un activateur est un bon exemple de coopération entre protéines dans la réalisation d'une fonction moléculaire élémentaire : ici le démarrage de la transcription. La régulation de l'opéron lactose (voir plus loin) se rattache à ce type de mécanisme.

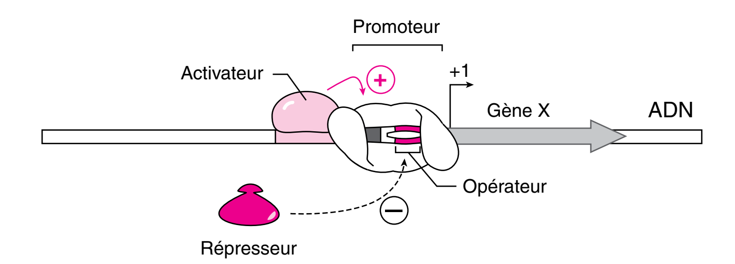


Figure 5.1 Schéma illustrant le rôle d'un activateur dans le recrutement d'une ARN polymérase à son promoteur.

Lorsque le répresseur occupe son site, il exclut par compétition l'enzyme de son promoteur.

Au cours de la transcription et avant même que l'ARN ne soit exporté hors du noyau de la cellule eucaryote pour être ensuite traduit (dans le cas d'un ARNm), il subit une série de modifications qui concourent à la production à partir du transcrit primaire d'un ARN dit mature, apte à assurer les fonctions auxquelles il est destiné. Dans l'ordre où elles interviennent on distingue pour les ARNm : l'addition d'une **coiffe** à l'extrémité 5', **l'épissage des introns** et l'addition d'une queue de **polyadénylation** à l'extrémité 3'.

L'exemple historique de l'opéron lactose de *E. Coli*

L'opéron lactose comporte trois gènes adjacents dénommés *lacZ, lacY* et *lacA (Fig 5.2)*. Le premier code la β-galactosidase, une enzyme qui hydrolyse le lactose en glucose et galactose, deux oses utilisés par la cellule pour ses besoins énergétiques ; *lacY* code une perméase membranaire qui importe le lactose dans la cellule ; *lacA* code une transacétylase qui purge la cellule des thiogalactosides toxiques introduits par la perméase. Un seul ARN messager polycistronique, transcrit à partir du promoteur situé en amont du gène *lacZ*, fournit par traduction les trois protéines. Elles sont produites seulement quand le lactose est disponible (on dit que le lactose agit comme un **inducteur**) et que le glucose est absent.

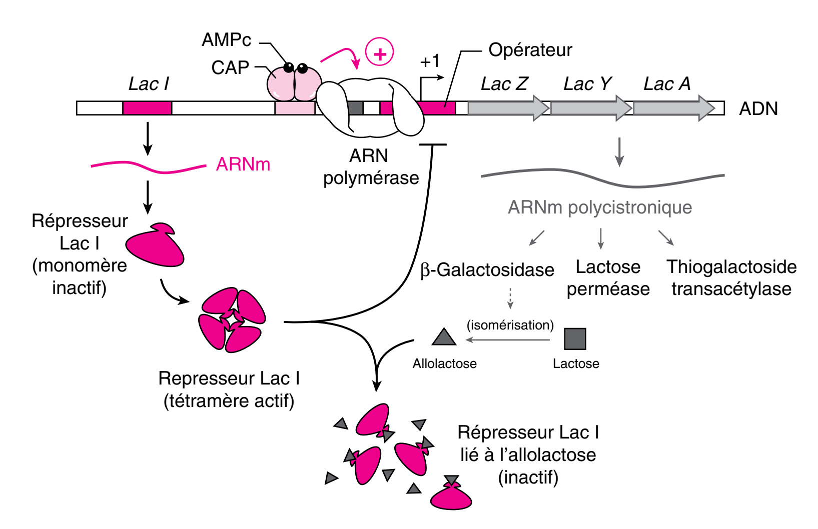


Figure 5.2 Positionnement des divers partenaires fonctionnels de l'opéron *lac*.

La transcription des gènes *lacZ*, *lacY* et *lacA* n'est effective que lorsque le lactose est présent c'est-à-dire lorsque le répresseur est inactif et l'opérateur inoccupé. La protéine CAP complexée à l'AMPc stimule le recrutement de la polymérase uniquement en l'absence du glucose.



En présence de glucose, la bactérie utilise préférentiellement cet ose pour ses besoins métaboliques en énergie et en composés carbonés. Si de plus le lactose est absent, le répresseur Lac réprimera fortement la transcription de l'opéron *lac*. Les quantités de β-galactosidase et des deux autres protéines sont produites à un niveau basal, soit environ 5 molécules par cellule. En présence du lactose et en l'absence du glucose, le répresseur devient inactif et la transcription de l'opéron *lac* est déréprimée. Mais la protéine CAP activera l'opéron *lac* seulement en l'absence du glucose. On peut dénombrer jusqu'à 5 000 molécules de β-galactosidase par cellule lorsque l'opéron est pleinement actif.

Le répresseur Lac et la protéine CAP sont deux exemples de protéines liant l'ADN en des sites spécifiques et uniques par leurs séquences en bases. Le site de l'opérateur *lac*, reconnu par le répresseur Lac, est une séquence partiellement palindromique de 21 paires de bases répétées et inversées, à double symétrie.

Chacune des deux sous-unités du répresseur comporte une structure spatiale très conservée par beaucoup de protéines liant l'ADN. Elle se compose de deux hélices α qui sont séparées par une boucle. Ce motif nommé hélice-tour-hélice s'attache à un demi-site de l'opérateur par l'une des hélices (l'hélice de reconnaissance) qui s'engage dans le grand sillon de l'ADN à l'aplomb de la séquence correspondante. La reconnaissance spécifique entre l'hélice protéique et la séquence en bases de l'ADN met en jeu des liaisons hydrogène et des interactions électrostatiques qui stabilisent la fixation de l'hélice au demi-site de l'opérateur *lac (Fig. 5.3)*.

La séquence de ce dernier chevauche la séquence du promoteur et lorsque le répresseur est présent, il empêche l'ARN polymérase de se fixer et inhibe la transcription de l'opéron *lac*. Le contact entre la protéine CAP et son site d'attachement à l'ADN obéit à ces mêmes principes structuraux. On note la présence d'un motif **hélice-tour-hélice** et des interactions basées sur des liaisons faibles et nombreuses entre les amino-acides de l'hélice engagée dans le grand sillon et les bases de l'ADN. L'action activatrice de la protéine CAP s'exerce par une interaction directe entre une région de la protéine (le **domaine activateur**) et le domaine carboxy-terminal de la polymérase. Le promoteur *lac* ne possède pas en effet d'élément UP (voir le Mini Manuel de *Biologie moléculaire*) et de ce fait, l'ARN polymérase s'y attache faiblement. On dit que le promoteur *lac* est un promoteur naturellement faible. Lorsque le site CAP est occupé par la protéine CAP, celle-ci recrute activement par son domaine d'interaction l'ARN polymérase et stimule ainsi la transcription.

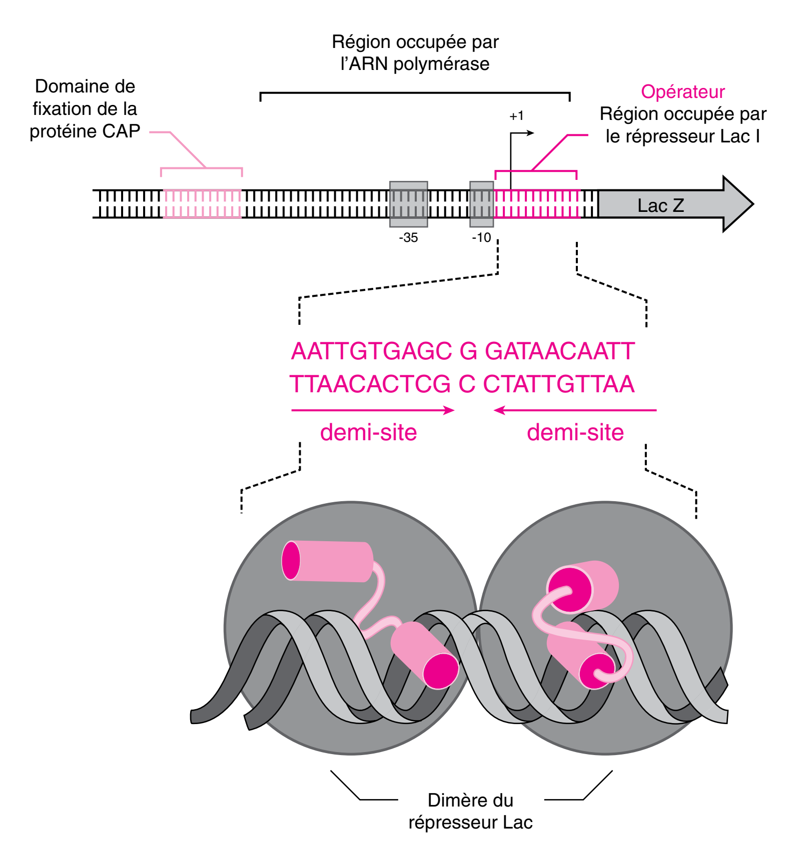


Figure 5.3 Organisation sommaire de la région de contrôle de l'opéron *lac* montrant les séquences d'ADN occupées par les protéines régulatrices et l'ARN polymérase. La structure palindromique en demi-sites de l'opéron lac est détaillée. L'interaction du répresseur Lac sous sa forme dimérique montre le positionnement des hélices dans le grand sillon de l'ADN.

Une même protéine régulatrice peut être répressive ou activatrice

La protéine « répresseur » du bactériophage λ, nommée le **répresseur λ**, est codée par le gène *cI (Fig. 5.4)*. Malgré son nom, cette protéine qui possède deux domaines, un domaine de liaison à l'ADN couvrant 17 pb et un domaine d'activité, peut activer ou réprimer la transcription. Elle existe habituellement sous la forme d'un dimère. Lorsqu'elle réprime, elle fonctionne comme le répresseur Lac et se fixe à un opérateur chevauchant le promoteur excluant ainsi l'ARN polymérase. Lorsqu'elle active la transcription, elle fonctionne comme la protéine CAP et recrute l'ARN polymérase.



La **protéine Cro** (Control of repressor and others things) est codée par le gène *cro*. Elle possède un seul domaine de liaison à l'ADN couvrant 17 pb et existe aussi sous la forme d'un dimère. Plus simple que le répresseur λ, elle réprime la transcription comme le fait le répresseur Lac.

Figure 5.4 Représentation schématique de l'action respective du répresseur λ et de la protéine Cro.

Noter les mécanismes d'autorégulation positive et négative du répresseur λ. Par l'occupation coopérative des opérateurs O_{R2} et O_{R3} il active ou réprime sa propre synthèse à partir de PRM.

Les isolateurs eucaryotes

Quand un activateur agit à longue distance (parfois à plusieurs milliers de kb d'un promoteur) il est inévitable que son action puisse s'exercer sur différents promoteurs présents dans l'intervalle. Cet inconvénient

semble, chez les eucaryotes, surmonté par la présence de séquences d'un type particulier, appelées **isolateurs**. Placés entre un site activateur (**enhancer**) ou un site répresseur (**silencer**) et un promoteur, l'isolateur inhibe l'activation ou la répression du gène par le **stimulateur** ou le **répresseur** en bloquant la communication entre les deux partenaires. Il n'empêche pas cependant l'activateur ou le répresseur d'agir sur un autre gène, ni le gène bloqué d'être activé ou inhibé par un autre stimulateur ou répresseur *(Fig. 5.5)*. Au total un isolateur en ciblant la protection d'un gène évite que des événements d'activation ou de répression sur de longues distances puissent altérer le fonctionnement du génome. En ce sens, les isolateurs sont des éléments plutôt spécifiques des génomes complexes et de grande taille.

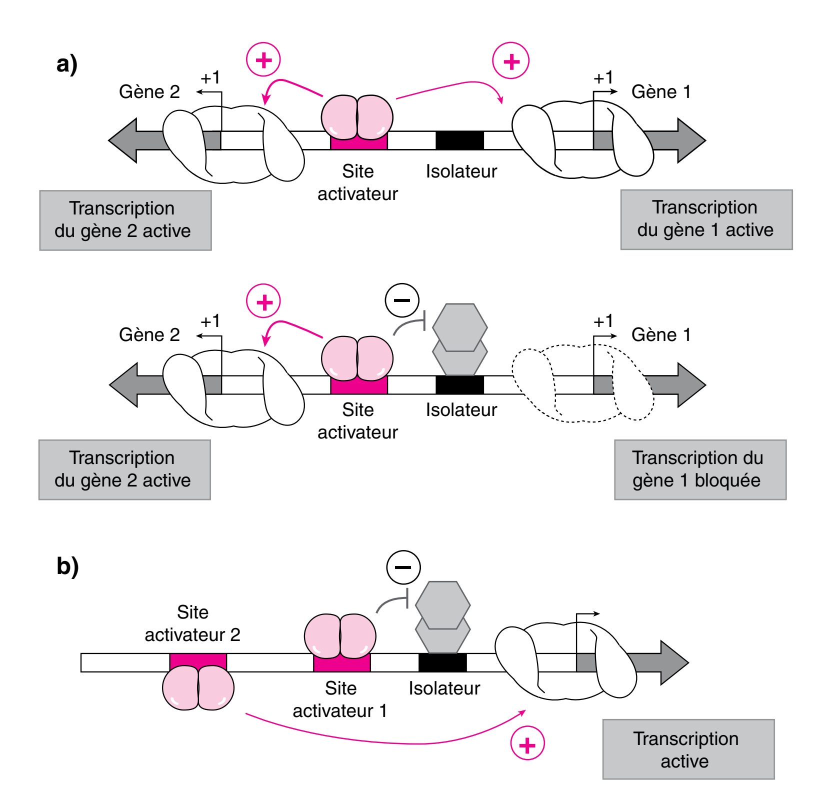


Figure 5.5 Illustration du mode d'action d'un isolateur.

a) L'isolateur occupé inhibe uniquement l'action de l'activateur sur le gène 1. Il n'interfère pas avec l'activation du gène 2 ; **b)** l'isolateur occupé n'interfère pas avec l'activation du gène 1 par un autre activateur (ici l'activateur 2).



a)
Protéine activatrice
Protéine répresseur
ADN
+1
Site activateur
Site répresseur
Compétition

b)
ADN
+1
Site activateur
Site répresseur
Inhibition

c)
ADN
+1
Répression directe

d)
Histone désacétylase
Répresseur
+1
ADN
Désacétylation des histones
Inhibition indirecte du promoteur par condensation de la chromatine

Figure 5.6 Exemple des principaux modes d'action des répresseurs eucaryotes.

a) Les sites de liaisons à l'ADN sont chevauchants et la présence du répresseur exclut celle de l'activateur ; **b)** les sites de liaisons sont distincts mais rapprochés, le répresseur peut inhiber par interaction directe l'effet de l'activateur ; **c)** le répresseur occupe son site et agit sur la machinerie transcriptionnelle par l'intermédiaire du complexe médiateur ; **d)** le répresseur, lié à son site, recrute une histone désacétylase qui en éliminant les groupes acétyles des queues d'histones, entraîne une modification de la structure des nucléosomes et de la chromatine, dont les effets sont inhibiteurs de la transcription.



La régulation transcriptionnelle à l'échelle de la chromatine

Contrairement au mode d'action le plus fréquemment rencontré chez les bactéries, les répresseurs et activateurs eucaryotes n'agissent pas, pour inhiber ou stimuler l'attachement des ARN polymérases aux promoteurs, par compétition et encombrement stérique du site promoteur. Ils peuvent cependant interagir avec les ARN polymérases en se liant à des sites adjacents à ceux des promoteurs, de façon alternative ou compétitive. Les activateurs et répresseurs eucaryotes recrutent fréquemment des protéines modificatrices de la structure chromatinienne *(Fig 5.6)*.

5.2 EXPRESSION DES GÈNES : L'ÉPISSAGE ALTERNATIF DES TRANSCRITS

Gènes morcelés et épissage des transcrits

Contrairement aux gènes bactériens dont les séquences codant les protéines sont continues, la plupart des gènes eucaryotes ont une séquence organisée en régions codantes (les **exons**) interrompues par des régions non codantes (les **introns**). On parle de gènes en mosaïques, ou de gènes morcelés ou éclatés. Cependant, comme les gènes bactériens, ils sont transcrits en une copie ARN continue, le transcrit primaire (appelé aussi **pré-ARNm**, dans le cas d'un précurseur d'ARNm). Comme le système cellulaire de traduction en protéines fonctionne d'une manière à lire en continu les codons successifs, les ARN messagers eucaryotes sont préalablement débarrassés de leurs introns. Le pré-ARNm est converti en ARN messager mature grâce à une machine moléculaire, le **splicéosome**, agissant avec une extrême précision sans perte d'un seul nucléotide ce qui aurait pour effet de rompre le cadre de lecture indispensable à l'obtention d'une protéine conforme à l'information génétique initiale. Fréquemment l'épissage des introns peut se produire de manière alternative (différentes combinaisons d'introns sont éliminées) et donner naissance pour un même gène à divers ARN messagers matures qui tous sont susceptibles de conduire à des protéines fonctionnelles mais distinctes. Les différentes protéines résultant de la transcription d'un seul gène sont des **isoformes protéiques**. Elles peuvent avoir des fonctions similaires, différentes ou antagonistes. Un gène peut donner parfois naissance alternativement à plusieurs transcrits. Certains s'avèrent parfois improductifs car ils portent un codon stop prématuré qui conduit à une protéine tronquée, non fonctionnelle et rapidement dégradée. Ce n'est pas toujours le cas. Les diverses modalités d'épissage connues sont souvent à l'origine d'une grande diversité de protéines fonctionnelles.



Exon 2
Exon 3
ADN
Promoteur
Intron 2
Exon 2 Intron 2 Exon 3
5′ A G G U A/G AGU A Pyr tract A G G 3′
Site d'épissage 5′
Point de branchement
Site d'épissage 3′

Exon 2 Intron 2 Exon 3
G G U A A G G
pré-ARNm

Première réaction de transestérification
5′ G G U A A G G 3′

Seconde réaction de transestérification
U G A A G G
5′ G 3′ 3′

Exon 2 Exon 3
5′ G G 3′ + Lasso excisé
Exons assemblés
U G A A G

Figure 5.7 Principales étapes de la réaction d'épissage des introns.

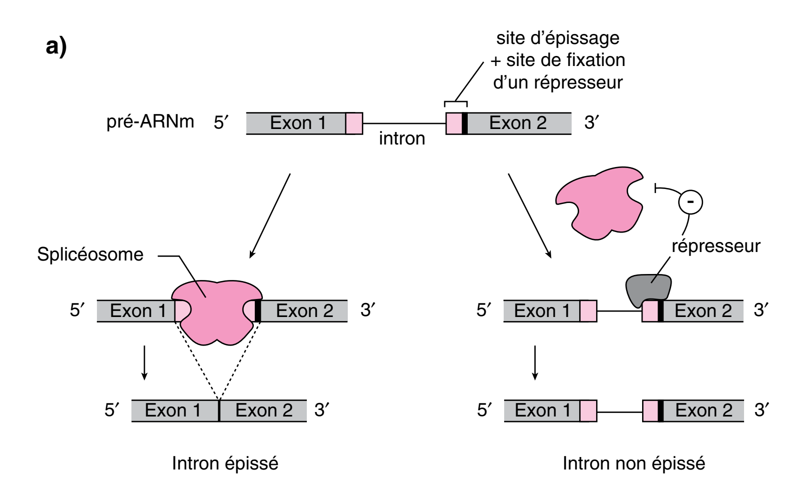
Noter dans le schéma du haut, les séquences consensus des sites d'épissage en bordures intron-exon et à l'intérieur de l'intron (voir aussi le Mini Manuel de *Biologie moléculaire*).



Épissage alternatif et régulation

Cas général

La conséquence admise de l'**épissage alternatif** est de diversifier très sensiblement l'expression des gènes eucaryotes, un phénomène qui est en accord avec la plus grande complexité des organismes qui les possèdent.



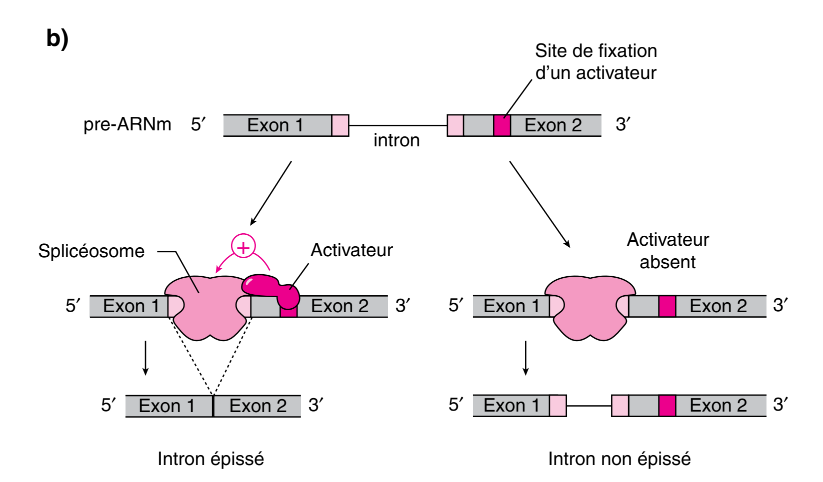


Figure 5.8 Principes généraux de la régulation de l'épissage alternatif montrant les rôles respectifs des sites d'épissage, du splicéosome, des sites régulateurs de l'épissage et des protéines régulatrices, activateurs ou répresseurs.

a) Le répresseur masque le site d'épissage et empêche l'action du splicéosome ; **b)** en occupant son site, l'activateur favorise le recrutement du splicéosome.



À l'intérieur de la séquence du transcrit primaire (**pré-ARNm**) au sein des introns, de courtes séquences consensus délimitent les frontières entre exons et introns *(Fig. 5.7)*. À l'extrémité 5' de l'intron la séquence s'appelle « **site d'épissage en 5'** » et à l'autre extrémité « **site d'épissage en 3'** ». Une troisième séquence consensus, le point de branchement, se trouve à l'intérieur de l'intron plutôt du côté 3' et elle est suivie d'une séquence riche en pyrimidines.

La régulation de l'épissage alternatif ressemble beaucoup à la régulation transcriptionnelle. L'attachement de la machinerie d'épissage à un site d'épissage particulier dépend en effet de l'affinité de ce site pour la machinerie et de l'action des protéines régulatrices de l'épissage. Un **site d'épissage** puissant favorisera un épissage constitutif. Mais un répresseur d'épissage peut entrer en compétition pour s'attacher à ce site et bloquer la machinerie d'épissage. Des séquences stimulatrices de l'épissage sont également présentes. Elles sont reconnues par des protéines régulatrices qui recrutent la machinerie d'épissage. Comme pour les protéines régulatrices de la transcription, ces protéines possèdent deux types de domaines, l'un qui s'attache à l'ARN et l'autre qui s'attache à la machinerie d'épissage *(Fig. 5.8)*.

Épissage alternatif et localisation cellulaire

La régulation de l'épissage alternatif des ARN messagers a pour conséquence principale d'accroître la diversité des protéines cellulaires et d'aboutir à la synthèse de protéines différentes selon les types cellulaires considérés. Un même transcrit pourra ainsi fournir deux ARNm matures codant pour deux protéines exerçant des fonctions de liaison vis-à-vis d'un ligand dans deux localisations cellulaires différentes *(Fig. 5.9)*.

5.3 EXPRESSION DES GÈNES : RÉGULATION TRADUCTIONNELLE

La séquence complète d'un ARNm a généralement une taille supérieure à la partie qui sera effectivement traduite en séquence protéique. Le terme « **cadre ouvert de lecture** » ou **séquence codante** (voir plus haut) désigne le segment d'ARNm qui est traduit *(Fig. 5.10)*. Chaque séquence codante est délimitée par un **codon de démarrage** (très souvent AUG) et **un codon de terminaison** de la traduction qui est l'un des trois codons d'arrêt, UAG, UGA ou UAA.

La plupart des ARNm eucaryotes comportent un seul cadre ouvert de lecture ; ils sont de ce fait dits **monocistroniques**. Les ARNm procaryotes comportent fréquemment plusieurs ORF et sont dits pour cette raison **polycistroniques**. Ils codent généralement chez les bactéries des protéines impliquées dans une même voie métabolique. Pour certains ARNm polycistroniques, le codon d'arrêt d'un ORF chevauche

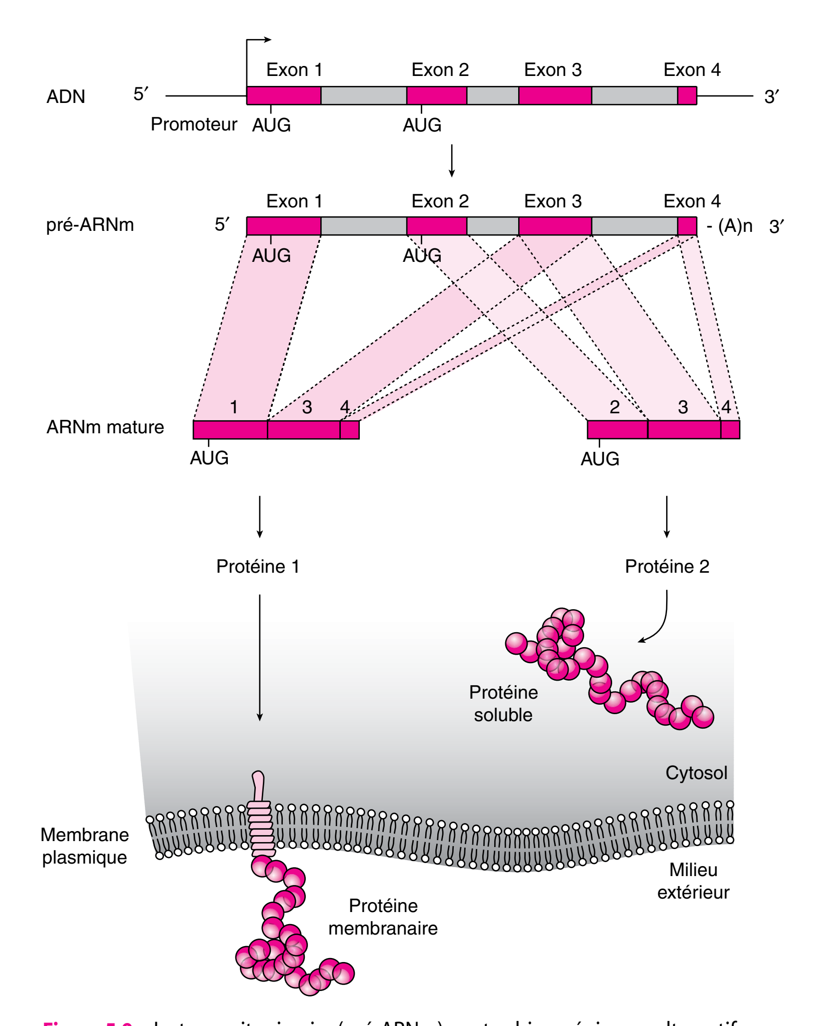


Figure 5.9 Le transcrit primaire (pré-ARNm) peut subir un épissage alternatif.

Les deux protéines obtenues possèdent le même domaine C-terminal codé par les exons 3 et 4 (même cadre de lecture) mais diffèrent dans leurs parties N-terminales qui sont codées respectivement par l'exon 1 ou l'exon 2. Les deuxs protéines n'ont pas la même localisation cellulaire. L'une est ancrée dans la membrane grâce à son domaine transmembranaire codé par l'exon 1 et exerce une fonction réceptrice à l'extérieur de la cellule, l'autre est strictement cytoplasmique et ne pourra s'associer qu'à un ligand intracellulaire.



le codon de démarrage de l'ORF suivant dans une séquence 5'-AUGA-3' contenant les deux codons. En règle générale les régions en amont 5' du codon AUG et en aval 3' du codon d'arrêt de l'ARNm sont cependant indispensables à la traduction. Elles sont différemment utilisées par les procaryotes et les eucaryotes pour le recrutement des ribosomes et des nombreuses protéines nécessaires à la traduction.

Les éléments de structure des ARN messagers eucaryotes influençant la traduction comportent *(Fig. 5.10)* :

- Les structures situées aux deux extrémités de l'ARNm : la structure en 5', constituée par la 7-méthyl-guanosine triphosphate, appelée aussi chapeau (« **capping structure** ») et la structure en 3', constituée d'une séquence additionnelle polyadénylée ou queue polyA. La communication entre ces deux structures est essentielle pour l'efficacité du démarrage de la traduction.
- Les séquences internes du messager assurant pour le démarrage de la traduction une entrée des ribosomes (ou **IRES**) indépendante de l'extrémité 5' (7-méthyl-guanosine triphosphate) du messager.
- Les petits cadres de lecture (uORFs), situés en amont du cadre principal (ou ORF, « open reading frame ») représenté par la séquence codante du gène, qui amoindrissent la traduction par avortement répété en raison de leur taille.

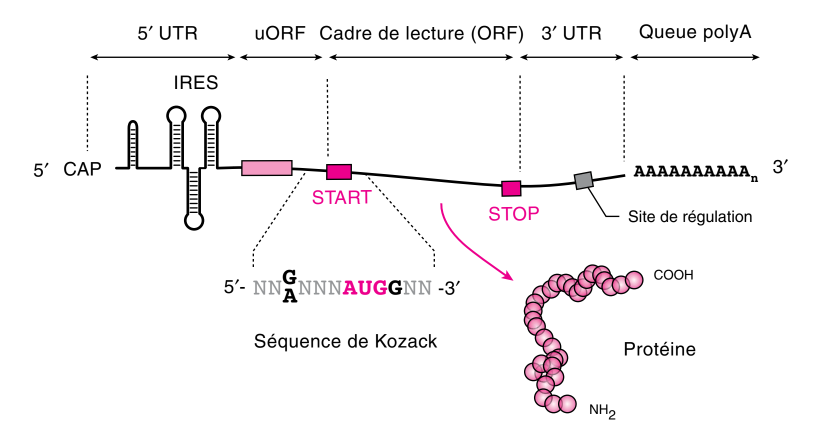


Figure 5.10 Structures des ARNm influençant la traduction.

L'ARNm eucaryote possède une extrémité 5' protégée (CAP) où se fixe le ribosome avant de balayer l'ARN jusqu'au codon de démarrage situé généralement au sein d'une séquence dite de Kozack stimulatrice du démarrage tout comme la queue polyadénylée située à son extrémité 3'. Noter la présence de séquence interne d'entrée des ribosomes (IRES) et de sites spécifiques de liaison de protéines ou d'ARN régulateurs. START : région contenant le codon de démarrage ; UTR : région non traduite.



- Les structures secondaires et tertiaires de l'ARNm, en forme d'épingles à cheveux ou de nœuds, qui bloquent normalement la traduction, mais peuvent faire partie des IRES et ainsi stimuler une traduction indépendante de l'extrémité 5'.
- Les sites de liaisons spécifiques de complexes protéiques qui sont des déterminants cruciaux de la traduction.

La plupart de ces éléments et mécanismes régulateurs sont en fait inhibiteurs de la traduction. Cela signifie que les ARNm d'une cellule eucaryote, loin d'être systématiquement traduits, une fois transcrits et maturés, se distribuent en deux ensembles d'inégale importance. Le premier, minoritaire, comporte à la fois un petit nombre d'ARNm activement traduits et les ARNm qui bien que non réprimés sont en attente de la disponibilité des facteurs de démarrage. Le deuxième, comporte tous les ARNm réprimés par l'un ou l'autre des mécanismes cités ci-dessus.

5.4 RÉGULATIONS ÉPIGÉNÉTIQUES DE L'EXPRESSION DES GÈNES

L'hérédité épigénétique repose sur la transmission d'une information qui n'est pas codée dans la séquence de l'ADN bien qu'elle soit transmise de cellule mère en cellule fille et de génération en génération. Des données expérimentales de plus en plus solides s'accumulent en faveur de l'existence d'un code particulier à cette hérédité. Il reposerait sur des modifications covalentes de l'ADN et sur un infini répertoire d'interactions établies entre la double hélice, les histones de la chromatine et de très nombreuses protéines régulatrices *(Fig. 5.11)*.

Modifications biochimiques des histones et de l'ADN : l'épigénomique

Les deux principaux composants biochimiques du code épigénétique héréditaire sont les mécanismes de méthylation-déméthylation de l'ADN au niveau notamment des îlots CpG et les modifications des queues d'histones par des mécanismes d'acétylation-désacétylation et de méthylation-déméthylation. Une nouvelle discipline, l'épigénomique se propose d'étudier ces phénomènes épigénétiques à l'échelle des génomes entiers. Elle recherche notamment à identifier les séquences de l'ADN qui dirigent les modifications épigénétiques et celles qui en sont la cible. Une attention spéciale est portée dans ces approches au rôle joué par les séquences répétées de type **LINE** (Long Interspersed Nucleic Element) et **SINE** (Short Intersperded Nucleic Element). Ces éléments sont largement représentés dans les génomes de vertébrés, notamment de mammifères, transmis fidèlement de génération en



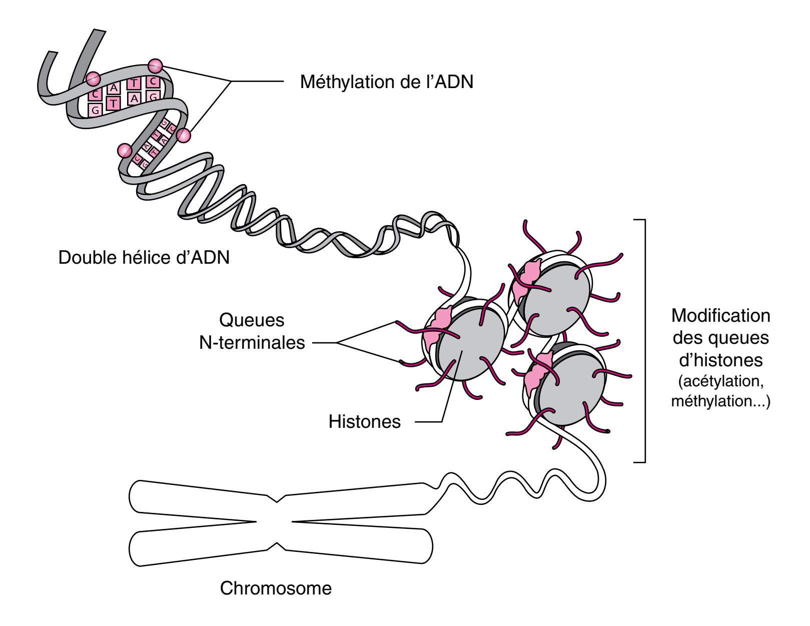


Figure 5.11 Les deux principaux composants du code épigénétique de la chromatine.

Les groupes méthyl (points roses) portés par certaines bases de l'ADN (cytosine principalement) contribue par leur présence à la répression des gènes. Des combinaisons de diverses protéines régulatrices peuvent se fixer aux queues protubérantes des histones et altérer l'expression de l'ADN proximal.

génération, et sont souvent mobiles dans les génomes sans que l'on ait encore bien compris qu'elles étaient réellement leurs fonctions génétiques.

Génétique et épigénétique des jumeaux monozygotiques

En raison de leur identité génétique parfaite – comme un clone composé de deux individus – les jumeaux monozygotiques sont depuis longtemps des modèles d'étude particulièrement appréciés pour tenter de distinguer le rôle respectif des gènes et de l'environnement dans la formation des phénotypes. Une étude récente portant sur les deux principales modifications épigénétiques (**méthylation de l'ADN** et **acétylation des histones**) accumulées au cours du temps par 40 paires de jumeaux monozygotiques, placés pour certains dans des environnements distincts, montre que 65 % des couples présentent des profils épigénomiques presque identiques, mais que 35 % manifestent des différences significatives. Une étonnante relation a été établie entre

l'âge des jumeaux et l'étendue des différences épigénétiques. Celles-ci sont également corrélées avec le temps de séparation qu'ont connu certains couples et leurs problèmes de santé. Afin d'établir la signification biologique de ces analyses, les empreintes méthylées de certaines séquences ont été soigneusement analysées. 52 % des séquences correspondent à des régions répétées et le reste à des gènes connus ou suspectés. Sont également impliqués des **îlots CpG** localisés dans des régions promotrices différemment méthylées, ce qui suggère qu'un effet sur l'expression des gènes s'est manifesté. Des analyses utilisant des microréseaux d'ADN confirment ces données. Les profils d'expression des jumeaux âgés de 3 ans sont quasiment identiques alors que ceux des jumeaux âgés de 50 ans diffèrent fortement. Il est connu que les modifications dans l'expression des gènes, dues à des marques épigénétiques accumulées au cours du temps, exercent une influence significative sur de nombreux types de maladies. Il reste à savoir si les différences de profils épigénétiques observées chez les jumeaux monozygotiques sont le résultat d'une accumulation de mutations touchant le maintien ou la transmission des marques épigénétiques ou si les facteurs environnementaux comme l'alimentation ou la pollution ont joué un rôle.

Modifications post-transcriptionnelles des transcrits par « éditing »

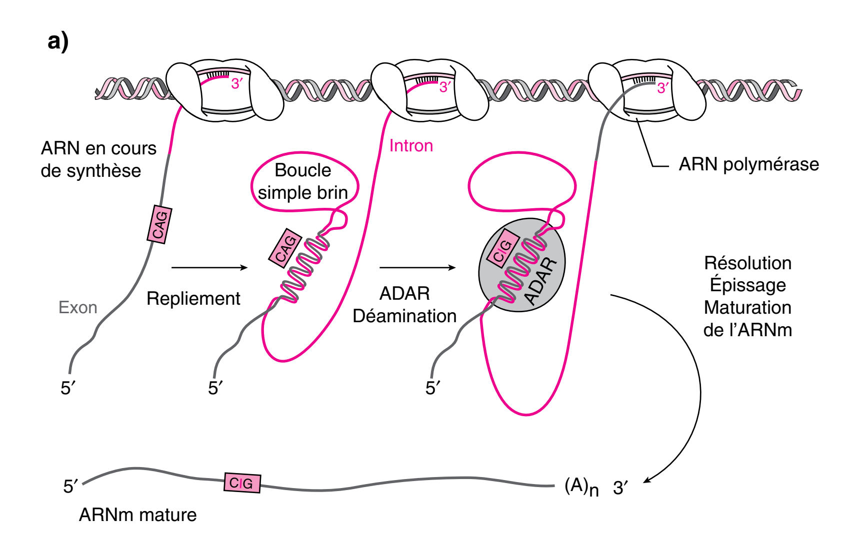
Il s'agit d'un mécanisme qui, à l'instar de l'épissage, modifie la séquence d'un ARN après sa transcription. La protéine qui en résulte est différente de celle qui est prédite par la séquence du gène.

Deux principales modalités sont connues : d'une part, la désamination d'une cytosine en uracile ou d'une adénine en inosine *(Fig. 5.12)* et d'autre part l'insertion ou la délétion d'uridine à l'aide d'un ARN guide. L'adénosine désaminase réductase (ADAR) transforme directement sur l'ARN, l'adénine en inosine. Au moment de la traduction, l'inosine sera reconnue comme une guanine et la séquence de la protéine en sera modifiée ainsi que sa fonction.

Empreintes parentales

Chez les organismes diploïdes, les cellules somatiques possèdent deux copies de génome et chaque **gène autosomal** est donc représenté par deux copies (deux allèles), chacune héritée de l'un des deux parents au moment de la fécondation. Pour la plupart des gènes autosomaux, les deux allèles généralement s'expriment. Cependant, certains gènes (< 1 %) sont porteurs d'une **empreinte parentale** qui détermine lequel des deux allèles pourra s'exprimer (*Fig. 5.13*). Par exemple le gène





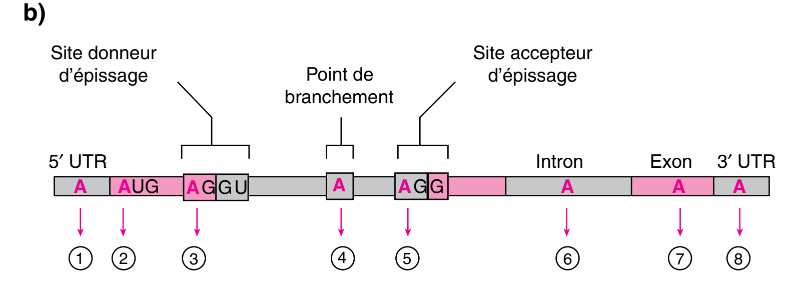


Figure 5.12 Mécanisme de l'« editing » de l'adénine en inosine (A→I) au niveau du pré-ARNm. Trois stades de l'action de l'ADAR (Adénosine déaminase réductase) sur le transcrit de l'ARN polymérase sont présentés.

(a) Le début du transcrit contient une adénosine modifiable au sein du codon CAG correspondant à la glutamine. La poursuite de la transcription dévoile une séquence intronique complémentaire contenant un site d'editing. Un appariement de bases s'établit et forme une hélice avec la région codante contenant le codon CAG. L'enzyme ADAR se fixe au duplex et désamine A en I ce qui a pour conséquence de donner un codon CIG (équivalent traductionnel de CGG). La glutamine sera remplacée par une arginine. Le duplex est ensuite dissocié par une hélicase et l'intron éliminé par le mécanisme habituel.

(b) Les différentes positions au long d'une molécule d'ARN où les adénosines peuvent subir un « editing ». Ces positions ne se rencontrent pas seulement dans les régions codantes des transcrits (2, 7) mais aussi dans les régions 3′ et 5′ UTR (1, 8), dans les sites d'épissage en bordures d'introns et d'exons (3, 5) et au site de branchement (4, 6). Un « editing » peut également modifier un codon stop (7).



codant IGF2 (Insulin-like growth factor) s'exprime uniquement *via* l'allèle paternel. L'addition d'une empreinte est un processus dynamique qui varie à chaque génération. L'empreinte peut s'effacer et se restaurer ce qui est en accord avec l'origine épigénétique (méthylation de l'ADN et modifications des histones) du processus.

Figure 5.13 Caractéristiques des gènes porteurs d'empreintes.

La figure présente schématiquement une paire d'allèles porteurs d'empreintes au niveau des îlots CpG et de séquences répétées (flèches). Les agrandissements soulignent les changements observés selon les allèles : condensation des nucléosomes consécutive à des désacétylations d'histones et des méthylations de l'ADN (allèle 1) ; ouverture de la chromatine suite à une acétylation des histones et une déméthylation de l'ADN (allèle 2). L'aptitude à la transcription de l'allèle 2 est indiquée par la fixation et l'action d'un complexe de transcription.



Dans la lignée germinale, l'empreinte est effacée puis elle est rétablie suivant le sexe de l'individu. Au cours de la formation des spermatozoïdes, une empreinte paternelle s'établit, alors que dans l'ovule une autre empreinte maternelle s'installe. Le processus d'effacement puis de reprogrammation s'avère indispensable afin que l'empreinte soit conforme au sexe de l'individu. Les gènes porteurs d'empreintes sont généralement regroupés en **clusters** et partagent des éléments régulateurs communs, comme des ARN non codants et des régions différentiellement méthylées. Ces dernières sont généralement riches en G + C, la cytosine étant souvent méthylée sur l'une des copies mais pas sur l'autre.

Les processus de mise en place des empreintes sont spécifiques aux mammifères placentaires et ils jouent un rôle dans le développement et la croissance embryonaires. Certains gènes de ce type participent au développement postnatal (allaitement et métabolisme). Plusieurs phénotypes majeurs dépendent des modalités de transmission ou de maintien d'empreintes parentales. Ainsi, à titre d'exemple, plusieurs maladies génétiques humaines sont associées à des empreintes anormales du locus 15q11 (la bande chromosomique 11 est située sur le **bras long** du chromosome 15). Cette région porte une empreinte épigénétique différente sur les chromosomes paternel et maternel, les deux empreintes étant cependant indispensables pour le développement normal de l'individu. Si l'une des empreintes n'est pas correctement transmise, en raison d'une délétion de la région touchant l'un des chromosomes ou si les deux chromosomes proviennent d'un seul des deux parents (disomie uniparentale), diverses maladies génétiques graves peuvent survenir. Ainsi, l'absence d'une empreinte paternelle se manifeste par le syndrome de Prader-Willi, caractérisé par une hypotonie, une obésité et un hypogonadisme ; l'absence de l'empreinte maternelle se manifeste par le syndrome d'Angelman caractérisé par une épilepsie, des tremblements et une expression faciale figée dans un sourire.

5.5 LES RÉSEAUX DE RÉGULATION DE L'EXPRESSION DES GÈNES

Analyse de l'expression à l'échelle des génomes : les outils de la génomique

La **génomique** étudie la totalité du génome d'un organisme à la différence de la génétique moléculaire qui s'attache plus spécialement à l'étude de gènes particuliers en accordant une attention à leurs divers rôles et fonctions. L'une de ses démarches principales consiste à établir la séquence complète des génomes étudiés (génomique structurale) et à procéder à l'analyse de l'expression dans sa globalité (génomique fonctionnelle) en fonction notamment les conditions environnementales.



Les outils privilégiés dans ces démarches reposent sur les analyses à haut débit, autorisées par les microréseaux d'ADN (puces à ADN) et la bioinformique. Des approches similaires sont pratiquées également avec l'ensemble des transcrits (transcriptome) d'une cellule ou d'un organisme, c'est la **transcriptomique**, ou avec les protéines (protéome), c'est la **protéomique**.

Au début de l'année 2005, plus de 1000 génomes viraux étaient séquencés, 250 génomes bactériens et près d'une trentaine de génomes eucaryotes. Beaucoup de ces génomes sont des organismes modèles (levure, drosophile, nématode, poisson zèbre, arabette, souris, rat, chimpanzé, homme) dont l'intérêt est grand à la fois sur les plans fondamental et appliqué.

Une branche très active de la génomique s'intéresse tout particulièrement à la comparaison des génomes et à leur évolution. Les données qui en résultent renouvellent et confirment à la fois les idées évolutionnistes proposées initialement par Darwin. Les relations génomiques entre des organismes vivants peuvent se résumer par une expression devenue célèbre : « tous parents et tous différents » qui intègre l'idée d'une commune origine (tous parents) et celle que mutation et sélection ont créé au fil du temps beaucoup de différences (tous différents). De nombreuses séquences génomiques sont conservées entre les espèces témoignant sans doute à la fois de l'avantage sélectif qu'elles représentent et de leur importance pour le fonctionnement du vivant. Mais toutes sont loin de coder des protéines, beaucoup sont transcrites en **ARN non codants**, ce qui a pour signification que le fonctionnement des organismes vivants semble étrangement dépendre dans une mesure aux limites encore inconnues de nombreux ARN dotés de fonctions régulatrices.

Les données fournies par les études de l'expression des génomes à grande échelle repose sur des approches utilisant les puces à ADN *(Fig. 5.14* et *Fig. 5.15)*. L'une des plus courantes aujourd'hui consiste à analyser de façon différentielle l'expression des gènes entre deux états physiologiques fonctionnels et à identifier ainsi ceux qui sont associés à la formation d'un phénotype particulier, qu'il soit biochimique ou morphologique *(Fig. 5.14)*. On notera que ce type d'analyse ne peut prendre en compte que les gènes dont les transcrits (ARNm polyadénylés) sont traduits en protéines et exclut les ARN non codants.

Un autre type d'analyse proche du premier consiste à identifier les séquences génomiques qui sont des sites de liaison pour les protéines exerçant une fonction régulatrice. Bien que la probabilité est faible que ces séquences soient situées à l'intérieur de gènes, la démarche est construite de telle manière qu'aucune éventualité ne soit exclue *(Fig. 5.15)*.

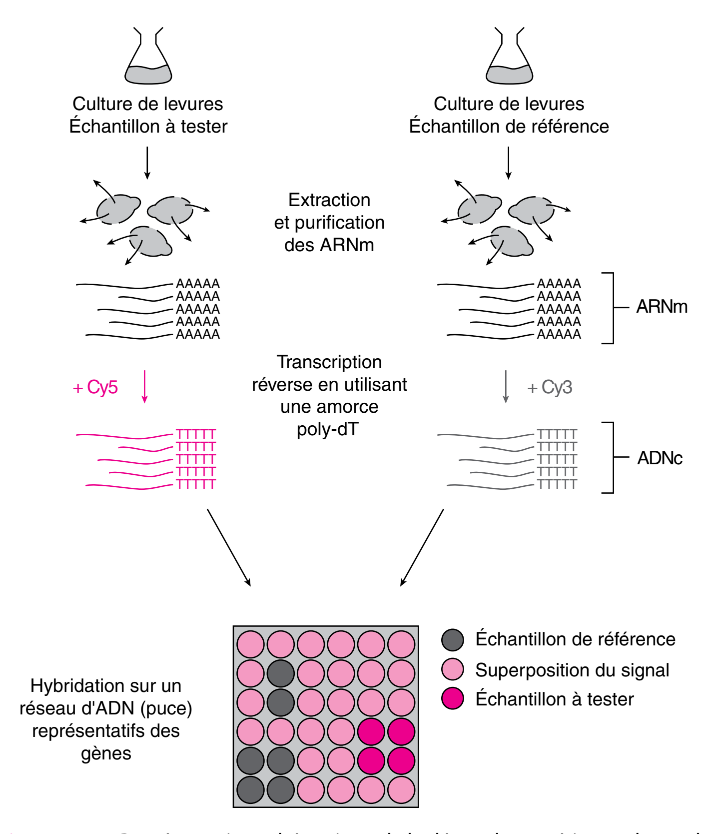
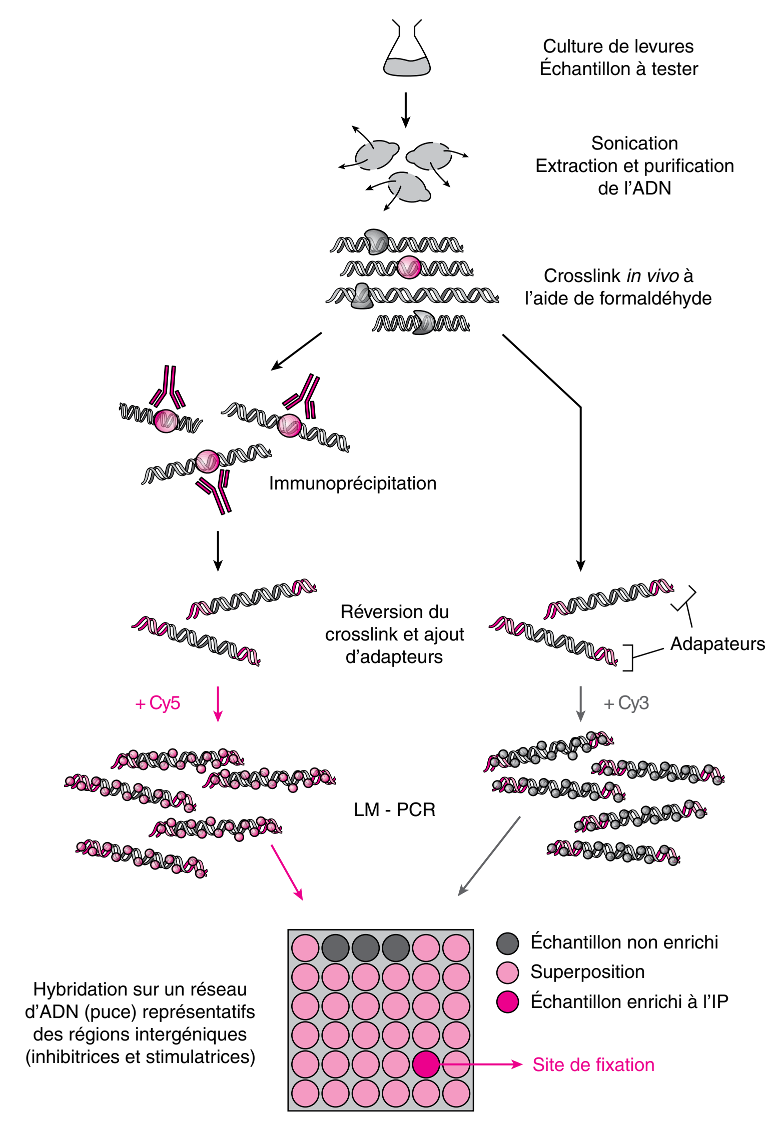


Figure 5.14 Représentation schématique de la démarche expérimentale conduisant à l'identification, à l'échelle du génome de la levure, des gènes dont l'expression est associée à un état physiologique (à un phénotype) particulier.

Les ARNm polyA, extraits des levures placées dans des conditions différentes de culture ou soumises à un stress biochimique ou physique, sont rétrotranscrits en ADN et marqués par des fluorochromes différents (cyanines 3 et 5). Ils sont ensuite hybridés sur un microréseau de cibles ADN représentatives des 6 200 gènes de la levure. L'analyse des fluorescences de chaque spot et le traitement informatique des données identifient les gènes dont l'expression varie en relation avec les conditions imposées initialement aux levures. En rouge, dans l'échantillon testé ; en noir dans l'échantillon de référence ; en rose, expression identique dans les deux échantillons.

Figure 5.15 Représentation schématique de la démarche expérimentale conduisant à l'identification, à l'échelle du génome de la levure, des séquences ADN liant des protéines régulatrices de gènes dont l'expression est associée à un état physiologique (à un phénotype) particulier. ➤



Les séquences ADN sont obtenues par une technique d'immunoprécipitation de l'ADN chromatinien après que la fixation des protéines de liaison à l'ADN ait été rendue irréversible par « crosslinking » à la formaldéhyde. Après addition d'adaptateurs et marquage fluorescent, les séquences d'intérêt et les séquences témoins sont hybridées et analysées grâce à un microréseau d'ADN représentatif des régions intergéniques (régulatrices) du génome de levure. La flèche rouge signale un « spot » identifiant une séquence ADN ayant liée une protéine régulatrice (inhibitrice ou stimulatrice de l'expression de gènes). Cette protéine est présente dans l'échantillon enrichi par immunoprécipitation (IP).



Les notions de hiérarchie et de réseaux de gènes

Ces notions peuvent être appréhendées expérimentalement par la mesure des interactions protéiques alliée aux données fournies par des « puces ADN » analysant le niveau des transcrits (*Fig. 5.16*). Cette démarche suppose que l'on dispose de données contrôlées et fiables. À partir du réseau maître qui donne l'ensemble des interactions établies entre les protéines A à F (en haut à gauche) certaines protéines peuvent être exclues selon les conditions cellulaires ou environnementales retenues (1 à 3) si l'on considère la perte d'expression de leurs gènes correspondants, constatée à l'aide d'une puce ADN (en haut à droite). On peut ainsi établir des sous-réseaux d'interactions protéiques spécifiques selon les conditions particulières retenues par l'expérimentateur. Ainsi, par une série d'expériences utilisant des puces à ADN, faciles d'accès aujourd'hui, on peut donc selon le type cellulaire ou les stades de développement considérés, établir une succession de réseaux

Protéines
A B C D E F
Protéines A B C D E F
Diagramme d'intéractions protéiques

Conditions
1 2 3
gènes A B C D E F
Données d'expression génique

Condition 1
Protéine B ⟷ Protéine C
Protéine A
Protéine D
Protéine E ⟷ Protéine F

Condition 2
Protéine B ⟷ Protéine C
Protéine A
Protéine D

Condition 3
Protéine B ⟷ Protéine C
Protéine A

Réseaux d'interactions protéiques conditionnels

Figure 5.16 Différents sous-réseaux d'interactions protéiques spécifiques de divers stades du développement d'un organisme ou d'une cellule peuvent se déduire d'un réseau protéique maître et de données acquises à l'aide de puces à ADN.



protéiques et géniques qui mettront en évidence les évolutions et les transitions adoptées par les réseaux de gènes qui contrôlent les processus développementaux.

De l'ensemble de ces démarches se dégage le concept considérant que les organismes fonctionnent dès la première cellule-œuf et les premiers stades de développement, sur la base de réseaux de gènes qui assureraient également leur adaptation permanente aux conditions environnementales. Différentes approches statistiques sont nécessaires pour dégager les relations régulatrices spécifiques (inhibitrices ou stimulatrices) existant entre les composants du réseau *(Fig. 5.17)*. De telles approches ont abouti à l'idée qu'à l'inverse d'une organisation de type hiérarchique avec des gènes « architectes » et des gènes « constructeurs » ou « réalisateurs » qui, à titre d'exemple, s'était dégagé des études génétiques sur la formation de l'œil, le fonctionnement du génome des organismes vivants s'apparente à un fonctionnement en réseaux de réseaux, dont la complexité est encore très loin d'être appréhendée.

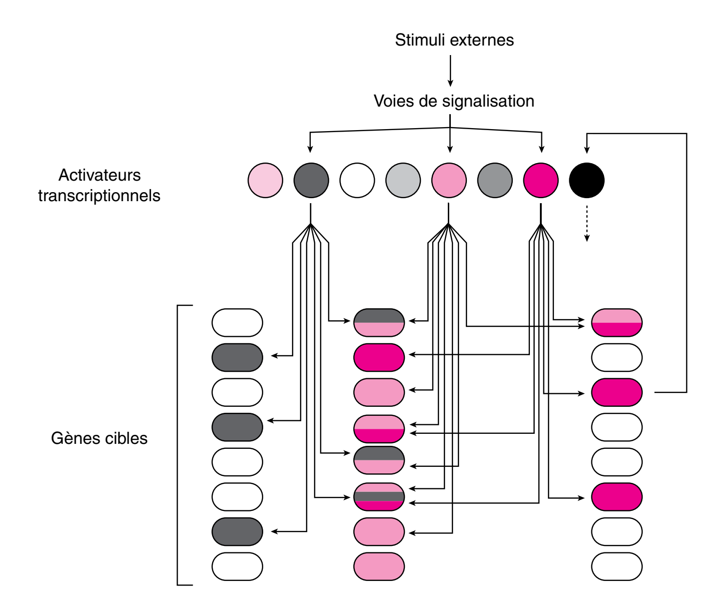


Figure 5.17 Schéma très simplifié d'un modèle de réseau régulateur de l'expression de gènes.

Les cercles colorés représentent des activateurs transcriptionnels. Les ovales rectangles représentent les gènes qui au sein du génome sont les cibles de ces activateurs. La couleur du rectangle indique quels sont les activateurs qui régulent l'expression du gène en réponse à un stimulus environnemental. Les flèches établissent la relation entre chaque activateur transcriptionnel et son gène cible. Un tel modèle peut être conçu comme un réseau individuel d'interactions géniques ou mieux encore comme une collection de réseaux régulateurs disposés eux-mêmes en réseaux, comme le suggère la flèche située la plus à droite du schéma.

Figure 5.18 Schéma récapitulatif spécifiant les flux d'information allant du génome au phénotype et indiquant les types de données utilisées pour décrire les processus biologiques intégrés.

De haut en bas, on trouve successivement, l'ADN (génomique) transcrit d'abord en ARN (transcriptomique) et traduit ensuite en protéines (protéomique) qui elles-mêmes peuvent catalyser des réactions qui agissent ou donnent naissance à des métabolites (métabolomique) parmi lesquels on trouve des glycoprotéines et des oligosaccharides (glycomique) et différents lipides (lipidomique). Beaucoup de ces composants peuvent être marqués et localisés au sein de la cellule ou de l'organisme (localisomique). Les processus responsables de la formation et de la modification des composants sont basés habituellement sur des interactions moléculaires (interactions entre protéines et ADN par exemple pour la transcription, ou entre protéines dans la traduction et les réactions enzymatiques). Finalement, les voies métaboliques comportent des réseaux intégrés, des cartes de flux (fluoxomique) qui dictent le phénotype cellulaire ou de l'organisme (phénomique).



Autres « omiques »

Les termes en « ome » (génome, transcriptome, protéome, métabolome, lipidome, glycome, etc.) et « omiques » dérivent du suffixe « -ome » dont l'utilisation a été généralisée pour désigner des champs nouveaux de la connaissance et de la recherche en biologie. Pour certains (génome et génomique par exemple) ces termes sont réellement fondés, ils sont plutôt spéculatifs pour d'autres. Le terme en « ome » désigne une collection d'objets (les gènes par exemple) dont il s'agit d'étudier la nature, l'origine, la fonction, etc., alors que le terme en « omique », qui possède à l'origine le statut d'un adjectif, désigne maintenant les sciences correspondant à ces objets avec leurs concepts, hypothèses et spéculations. Compte tenu qu'il s'agit dans tous les cas d'acquérir et de manipuler de grands ensembles de données parfois complexes, l'intervention de l'informatique est une constante de ces champs nouveaux de la biologie.

Les sciences « omiques » se proposent de fournir des descriptions détaillées de tous les composants et interactions qui prennent place dans une cellule ou un organisme vivant. Ces données peuvent être classifiées en trois catégories : les composants, les interactions, et les données relatives aux états fonctionnels (*Fig. 5.18*). Les données relatives aux composants détaillent le contenu moléculaire de la cellule ou du système étudié ; les données relatives aux interactions spécifient les liens existant entre les composants moléculaires et les données relatives aux états fonctionnels fournissent une lecture intégrée de tous les types de données « omiques » en révélant un phénotype global de la cellule ou du système.

5.6 LA GÉNOMIQUE EN SANTÉ HUMAINE ET POUR LA SÉLECTION ANIMALE

La génomique en santé humaine

Découvrir les variations génétiques héréditaires qui sont à l'origine des pathologies humaines, constitue aujourd'hui l'un des défis scientifiques majeurs de la biologie et de la génétique. Le séquençage complet en 2001 du génome humain a représenté, à cet égard, un progrès considérable, mais l'exploit doit être atténué par le fait qu'il s'agissait alors de la seule séquence génomique haploïde. Aucune indication ne mettait en évidence la présence des variations de séquence qui sont, chez les humains comme chez tous les autres êtres vivants, à l'origine pour l'essentiel des nombreuses différences phénotypiques. Depuis lors, de plus en plus d'études ont été conduites pour découvrir les liens de causalité existant entre les variations de la séquence génomique et la susceptibilité des humains vis-à-vis des maladies qui les affectent.



Les variations structurelles du génome humain

Ces études conduisent à distinguer parmi les polymorphismes génétiques humains identifiés, ceux dits « communs » présentant une fréquence allélique d'au moins 1 % au sein de la population et ceux dits « rares » dont la fréquence allélique est inférieure à 1 %. D'un point de vue structural, on peut distinguer deux types de variations génétiques : les SNP (pour *Single Nucleotide Polymorphism*) et les polymorphismes dits structuraux (Tableau 5.1).

TABLEAU 5.1 LES DIFFÉRENTES CLASSES DE POLYMORPHISMES GÉNÉTIQUES HUMAINS.

Type de variation	Exemple
SNP (polymorphisme d'un seul nucléotide)	ATGGCCTTTACCCCGACGATCAGGAT ATGGCCTTTACTCCGACGATCAGGAT
Indel (insertion/délétion)	ATGGCCTTTACCCCGACGATCAGGAT ATGGCCTTTACCCC---GATCAGGAT
Substitutions de blocs	ATGGCCTTTACCCCGACGATCAGGAT ATGGCCTTTAGAGTGACGATCAGGAT
Inversion	ATGGCCTTTACCCCGACGATCAGGAT ATGGCCTTCGTCGGGGTAATCAGGAT
Variation du nombre de copies	ATGGCCTTTAGCCTTTACCCCGACGATCAGGAT ATGGCCTTTA-------CCCCGACGATCAGGAT

Les SNP sont des variations de séquences ADN où un simple nucléotide (A, T, G ou C) est modifié. Les autres variations dites structurelles comprennent : les « Indel » (pour insertion-délétion) qui se rencontrent lorsque une ou quelques bases sont présentes dans un génome et absentes dans les autres. Composées généralement de quelques bases elles peuvent parfois aller jusqu'à 80 kb. Les substitutions de blocs désignent des séquences de nucléotides adjacents qui varient entre deux génomes. Les inversions sont observées lorsque l'ordre des paires de bases s'inverse dans une section définie d'un chromosome (on connaît une inversion d'une longueur de 900 kb dans le chromosome 17 humain, chez 20 % des individus ayant un ancêtre nord européen). Les variations du nombre de copies (ou CNV pour *Copy Number Variation*) désignent des séquences identiques ou presque qui sont répétées chez certains individus mais pas chez d'autres. Le nombre de répétitions varie et ces dernières peuvent atteindre parfois une taille élevée de l'ordre de 2 Mb.

L'origine des polymorphismes génétiques doit être recherchée dans les évènements aléatoires non corrigés survenant lors de la duplication de l'ADN et de la recombinaison chromosomique concomitante de la mitose et de la méiose (voir chapitre 2). La plupart des variations génétiques, au sein de la population humaine sont ainsi dues au hasard

de la reproduction et ne contribuent probablement pas à la diversité phénotypique. On dit qu'elles sont neutres. Cependant pour une minorité d'entre elles, ces variations peuvent être à l'origine d'une modification de phénotype. C'est notamment le cas lorsqu'un SNP se situe dans une séquence codante ou à proximité d'un élément régulateur. Ainsi, lorsqu'un polymorphisme modifie un gène, ou une structure génétique, impliqué dans une pathologie, celui-ci peut participer à la susceptibilité ou à la résistance d'un individu à cette pathologie.

Les SNP

Ils sont les plus fréquents et l'on estime à près de 11 millions le nombre de SNP total au sein du génome humain (7 millions ayant une fréquence supérieure à 5 %, le reste une fréquence comprise entre 1 et 5 %). L'analyse de ces SNP au sein des premiers génomes individuels entièrement séquencés montre que la plupart d'entre eux sont répertoriés dans la base de données générale des SNP (*Tableau 5.2*). Il a été montré que les allèles des SNP présents dans une même région génomique sont souvent corrélés les uns aux autres. Cette corrélation, désignée par le terme de « déséquilibre de liaison », varie à la fois de façon complexe et imprévisible, au sein des génomes et entre les différentes populations.

TABLEAU 5.2 POLYMORPHISMES DE TYPE SNP IDENTIFIÉS DANS QUATRE GÉNOMES REMARQUABLES.

Génome	Nb de SNP	% de présence dans la base générale des SNP
J. Craig Venter	3 213 401	91, 0 %
James D. Watson	3 322 093	81, 7 %
Chinois ethnie Han	3 074 097	86, 4 %
Africain ethnie Yoruban	4 139 196	73, 6 %

C. Venter et J. Watson sont deux célèbres scientifiques parmi les premiers humains dont le génome a été entièrement séquencé. Ils sont de type caucasien. Les deux autres individus sont des anonymes de types asiatique et africain. Les trois premiers génomes possèdent un nombre similaire de SNP bien représentés dans la base actuelle de données des SNP humains. Au total 5, 2 millions de SNP différents sont identifiés entre les trois génomes. Lorsqu'un plus grand nombre de génomes individuels sera séquencé, 11 millions de SNP devraient être identifiés. Le génome africain possède un plus grand nombre de SNP, plutôt nouveaux vis-à-vis de la base de données. Cependant dans l'ensemble, il apparaît que lorsque deux génomes humains sont comparés, la majorité des bases qui diffèrent appartient à la catégorie des variations communes. La base de donnée générale (dbSNP) est accessible librement à l'adresse suivante : http://www.ncbi.nlm.nih.gov/projects/SNP/



Les polymorphismes structuraux

Pris dans leur ensemble, les polymorphismes structuraux représentent entre 9 et 25 Mb (soit jusqu'à près de 1 % du génome) ce qui suggère un rôle très significatif de ces variations génétiques en matière d'évolution du génome. (*Tableau 5.3*).

Tableau 5.3 Polymorphismes structuraux identifiés dans le génome de Craig Venter.

Polymorphismes structuraux	Nombre	Longueur (en pb)
Substitution de bloc	55 823	2 - 206
Indels (hétérozygotes)	851 575	1 – 82 711
Inversions	90	7 – 670 345
Nombre de copies	62	8 855 – 1 925 949

Le génome d'un individu caucasien (C. Venter) contient au total environ 4,1 millions de variations génétiques dont 22 % sont des polymorphismes structuraux représentant 74 % de toutes les bases variantes. Cependant, de nombreux polymorphismes structuraux restent vraisemblablement à découvrir avec l'augmentation du nombre de génomes individuels séquencés. Comme pour les SNP, on s'attend à ce que la majorité de ces variations soit commune à beaucoup d'humains.

Contribution des divers polymorphismes génétiques aux phénotypes chez l'homme

Chez les humains des centaines de phénotypes complexes déterminent l'allure extérieure, le comportement, mais aussi la propension à développer certaines maladies. On désigne sous le terme de « trait complexe » le phénotype qui, selon les individus examinés, présente divers degrés de variation anatomique, physiologique ou biochimique formant ainsi une sorte d'échelle de distribution continue. On considère que de tels phénotypes sont les produits de l'action indépendante de nombreux gènes (ou mieux de leurs allèles), mais aussi de facteurs environnementaux ainsi que de subtiles interactions entre gènes (allèles) et facteurs de l'environnement. Près de 4 millions de polymorphismes génétiques sont présents en moyenne simultanément chez un individu (*Tableaux 5.2 et 5.3*). La difficulté consiste à établir lesquels parmi ces polymorphismes, s'avèrent responsables de la transmission héréditaire des phénotypes, notamment ceux qui prédisposent à développer une pathologie (cancers, maladies cardio-vasculaires, maladies neurodégénératives, etc.). Deux théories opposées

(théories 1 et 2, voir l'exemple ci-dessous) sont alternativement proposées à ce sujet. La première affirme que les traits complexes sont principalement dus à un grand nombre de polymorphismes communs, chacun d'entre eux étant porteur d'un effet modeste. À l'opposé, la seconde propose que les traits complexes résultent d'une somme d'effets majeurs portés par un petit nombre de polymorphismes rares.

Les études d'associations entre phénotypes et génotypes.

Avec la disponibilité du séquençage complet du génome humain et la découverte de millions de SNP au sein des génomes, de nouveaux outils de génotypage ont été rapidement développés. Ces outils à haut-débit, appelées « puces ADN », permettent de génotyper simultanément plusieurs centaines de milliers, voire plusieurs millions de SNP répartis sur l'ensemble du génome. La disponibilité de ces nouvelles ressources a rendu possible la recherche, à l'échelle du génome dans son ensemble, d'une association statistique entre un phénotype (par exemple une pathologie) et certains polymorphismes génétiques communs ; cette approche est appelée étude d'association à grande échelle.

D'une manière générale, ces études d'association à grande échelle exploitent les différences de polymorphismes génétiques qui sont présents chez un grand nombre d'individus, souvent plusieurs milliers, atteints par une pathologie (les cas) comparativement à ceux présents dans un nombre similaire de patients sains (les contrôles). On parle d'études d'association « cas-contrôles » ou « cas-témoins ». L'objectif ici est de rechercher s'il existe certains polymorphismes pouvant être statistiquement associés à la maladie étudiée, et influencer la susceptibilité ou la résistance d'un individu au développement de cette pathologie. La majorité des polymorphismes mis en évidence par ces études d'associations à grande échelle ne se sont pas révélés être des mutations « causales » de la maladie, et n'expliquent généralement qu'une petite partie de la maladie étudiée (voir exemple ci-dessous). Ce type d'étude reste cependant très utile car elle permet de cibler des régions du génome associées à une pathologie et permet de restreindre ainsi les régions à explorer à la recherche de gènes candidats et de mutations causales. Les SNP identifiés jouent alors un rôle de marqueurs moléculaires.

Ainsi, si plusieurs SNP participant à une même voie de signalisation sont identifiés, il est vraisemblable que cette voie de signalisation participe au développement de la maladie. De même si plusieurs pathologies sont statistiquement associées à un locus particulier, cela



suggère une base génétique commune à ces maladies. Plusieurs SNP d'un locus du chromosome 8 ont été associés au cancer de la prostate, du colon et du sein, suggérant que ce locus est impliqué dans le processus de carcinogenèse. Des études supplémentaires ont permis de confirmer ces résultats et d'identifier la mutation causale, située dans un stimulateur de la transcription.

Outre une meilleure compréhension des bases moléculaires et des gènes impliqués dans la résistance ou la susceptibilité à certaines maladies, les résultats de ces études rendent possible le testage génétique. Ainsi il devient possible pour un patient de connaître le risque (relatif) qu'il présente vis-à-vis d'une pathologie ou encore de pouvoir recevoir une information personnalisée sur la réponse de son organisme à un composé pharmaceutique. Cette dernière application, appelée pharmacogénomique, est un enjeu important en terme de santé humaine. À titre d'exemple, des études d'associations ont permis de montrer que des polymorphismes du gène CYP2C9 sont associés au métabolisme de la warfarine, un anticoagulant utilisé pour prévenir la formation de caillots dans les vaisseaux sanguins. En fonction de la forme allélique qu'ils portent, certains patients vont assimiler plus ou moins rapidement ce composé, ce qui risque d'engendrer une hémorragie ; la connaissance de leurs variations génétiques permettra donc de personnaliser la dose de warfarine à donner dans les traitements afin d'éviter ces effets secondaires.

Comme le montre la figure 5.19, ce type d'approche à littéralement explosé au cours des dernières années (*Fig. 5.19a*). En 2009, près de 250 études d'association à grande échelle ont été réalisées chez l'homme. Mi-2010, 599 études de ce type ont été publiées et ont permis d'identifier 2 905 SNP associés à près de 150 pathologies humaines ! Par ailleurs, la réduction du coût de ces études, liées à des outils de plus en plus performants permet d'analyser des cohortes de patients de plus en plus importantes et d'organiser notamment des phases de validations indépendantes, indispensables pour valider les associations observées au cours d'une seule étude (*Fig. 5.19b*).

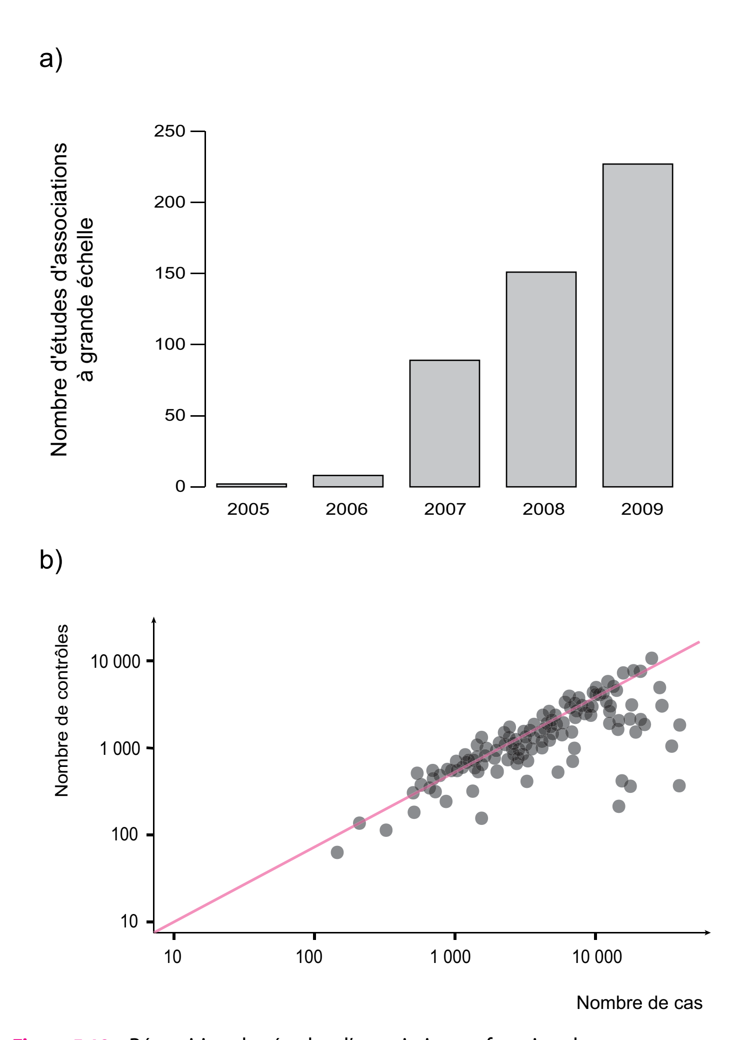


Figure 5.19 Répartition des études d'association en fonction du temps.

a. Les premières études d'associations à grande échelle datent de 2005 ; depuis leur nombre ne cesse d'augmenter d'année en année, notamment en raison d'outils de génotypage de plus en plus performants et de moins en moins coûteux. Les approches statistiques mises en place pour déterminer les associations sont également de plus en plus performantes.

b. Il est important de noter que ces études nécessitent des cohortes de grande taille pour être efficace. Une des limites principales de ces approches reste le trop faible nombre de patients participant à ces études. Malgré tout, la taille des cohortes utilisées dans les études d'association est de plus en plus importante, ce qui augmente beaucoup la puissance de ce type d'étude. L'ensemble des études d'associations réalisées chez l'homme est centralisée et accessible librement à l'adresse suivante : http://www.genome.gov/gwastudies/



Polymorphismes génétiques du diabète de type 2 : un exemple illustrant les deux théories à propos des traits complexes

Au total, dix-huit polymorphismes génétiques augmentant le risque de développer un diabète de type 2 chez des individus caucasiens ont été mis en évidence au cours d'études d'association à grande échelle. Si quatre d'entre eux avaient déjà été suspectés au cours d'études ciblant des gènes candidats, les quatorze autres SNP n'avaient jamais été identifiés dans la susceptibilité génétique à cette pathologie, ce qui a permis de cibler de nouveaux gènes, et de nouvelles voies de signalisation jusqu'alors ignorées. À titre d'exemple, l'identification d'un polymorphisme à l'intérieur d'un récepteur à la mélatonine (*MTNR1B*) a permis de montrer que la voie de signalisation contrôlant le cycle circadien joue aussi un rôle dans le diabète de type 2.

Classiquement, la force de l'association calculée lors d'études de type « cas-témoins » est mesurée par les *odds ratios* ; cette notion se définit par le rapport des probabilités qu'un polymorphisme génétique existant dans la population de patients atteints soit également présent dans la population contrôle. Si l'*odds ratio* est significativement supérieur à un, alors le polymorphisme est associé à la maladie. Dans le cas du diabète de type 2, les fréquences alléliques minoritaires (MAF) et les *odds ratios* calculés pour les 18 SNP intervenant dans l'augmentation du risque, sont similaires aux valeurs obtenues lors d'études d'associations conduites sur d'autres maladies. Les MAF s'échelonnent de 0,073 à 0,50 et les *odds ratios* vont de 1,05 à 1,15 sauf pour le gène *TCF7L2* qui possède un *odds ratio* de 1,37.

Au total, ces 18 polymorphismes expliquent moins de 4 % de la part totale du phénotype, et une faible fraction de son héritabilité. Ceci a pour conséquence qu'il reste pour cette pathologie de nombreux gènes à découvrir. Si l'on suppose que les polymorphismes génétiques encore inconnus présentent des MAF et des *odds ratios* similaires à ceux déjà identifiés et une héritabilité de 40 %, il resterait plus de 800 polymorphismes génétiques à découvrir (théorie 1) ! Si, au contraire, on considère qu'il s'agit de mutations rares, avec des MAF 10 fois plus petites que celles déjà identifiées (donc s'échelonnant de 0, 0073 à 0,05) et des *odds ratios* 10 fois plus grands (de 1,63 à 4,05) alors environ 85 nouveaux polymorphismes devraient être impliqués dans le diabète de type 2 (théorie 2).

Polymorphismes génétiques de type SNP et longévité humaine

Dans les pays industrialisés, une personne sur 6 000 est centenaire, et ceux que l'on appelle les supercentenaires - ayant plus de 110 ans – représentent un individu sur 7 millions. Il faut noter que 90 % des centenaires n'ont pas de handicap sur le plan de la santé avant l'âge de

93 ans. Dans une étude portant sur l'analyse complète du génome de 1 055 centenaires comparativement à 1 267 sujets contrôles, des chercheurs ont bâti un modèle prédictif prenant en compte 150 SNP. Il en ressort que lorsque l'on soumet les informations génétiques (c'est-à-dire les génotypes des 150 SNP) d'un individu à ce modèle mathématique, il est possible de déterminer correctement dans 77 % des cas sa prédisposition à devenir centenaire, ce qui atteste fortement du caractère héréditaire de la longévité. Cependant la présence de certains facteurs de risque (cancer, maladies cardio-vasculaires, etc.…) amène à moduler selon des groupes de risques les voies permettant d'atteindre cette extrême longévité. La surprise vient du fait que la présence de polymorphismes génétiques associés à des maladies ne met pas en évidence de différence marquée entre les centenaires et les sujets contrôles. Cela signifie que ce n'est pas l'absence de prédisposition pour des maladies qui fait vivre longtemps mais bien le fait d'être porteurs de mutations favorisant la longévité. Cela implique aussi que si l'on veut calculer le risque de développer une maladie, il ne faut pas seulement s'appuyer sur la présence de polymorphismes génétiques qui lui sont associés mais tenir compte du patrimoine génétique global.

La génomique pour la sélection animale

Contrairement à ce que l'on observe en matière de santé humaine où les informations tirées de la connaissance du génome humain sont, malgré de solides espoirs, encore peu exploitées, il est vraisemblable que la génomique animale va modifier rapidement et fortement les pratiques de l'élevage.

L'utilisation de la génomique, pour les animaux de ferme, n'apparaît pas pourtant *a priori* très différente de celle qui est envisagée en matière de santé humaine. Pour tous les cheptels, l'information génomique doit en effet conduire, sous peine d'inefficacité, à l'amélioration de leur état général et de leur productivité. Dans son principe il s'agira, tout d'abord d'identifier le ou les gènes d'intérêt, puis de sélectionner préférentiellement les allèles favorables au détriment des allèles défavorables. Pour cela il n'est pas réellement nécessaire d'identifier les gènes ou les mécanismes moléculaires sous-jacents, par exemple du rendement laitier, de la qualité de la viande ou de la fertilité. Il suffit de corréler quantitativement l'amélioration du trait avec certains polymorphismes génétiques puis d'établir ensuite, un programme de sélection qui accroîtra la représentation de ces mutations au sein de la population constituant le cheptel.

Ainsi à partir d'une première ébauche d'assemblage de la séquence du génome bovin, les chercheurs et les industriels ont créé une « puce



ADN » permettant de génotyper plus de 50 000 SNPs. Disponible depuis janvier 2008, cet outil a permis dès le mois de juillet suivant à l'industrie laitière d'améliorer ses programmes de sélection. La disponibilité croissante de banques d'ADN, très représentatives des populations bovines étudiées, ainsi que les nouvelles méthodes de séquençage à haut débit ont permis d'identifier rapidement un grand nombre de SNP exploitables pour créer ce type d'outils. Des outils similaires sont disponibles pour d'autres espèces animales telles que le chien, le cheval, le mouton, le porc ou encore le poulet.

Une autre différence capitale qui distingue les applications de la génomique chez l'homme et l'animal, réside dans le fait que l'information moléculaire peut s'appuyer, chez les principales espèces d'animaux de ferme, sur de grandes bases de données phénotypiques et généalogiques déjà en place au moment où les séquences génomiques deviennent disponibles. Il devient alors possible de rechercher des marqueurs génétiques associés aux phénotypes d'intérêt pour organiser une sélection basée sur cette information génomique.

Il n'en demeure pas moins que le grand intérêt de la sélection génomique réside dans le fait qu'elle est moins traumatisante, à la fois pour le public et pour les autorités de contrôle, que ne le sont l'obtention d'animaux transgéniques ou d'animaux clonés, tout en permettant d'accélérer le progrès génétique. Si la sélection génomique tient ses premières promesses, la santé animale, le bien-être animal et la productivité de l'élevage s'en trouveront profondément changés.



POINTS CLEFS

- Les mécanismes de la transcription reposent sur l'activité d'une famille d'enzymes, les ARN polymérases, qui réalisent le même type de réaction (une synthèse d'ARN à partir d'une matrice d'ADN) quelle que soit la cellule procaryote ou eucaryote considérée. Les ARN polymérases se fixent à des régions de l'ADN désignées sous le terme de promoteurs**,** souvent avec l'aide de protéines additionnelles. Les ARN polymérases adoptent une structure tridimentionnelle en forme de pince ce qui leur permet d'enserrer le brin d'ADN (orienté dans le sens 3′ → 5′) et en coulissant de le transcrire en un ARN orienté dans le sens 5′ → 3′.
- Au cours de la transcription et avant même que l'ARN ne soit exporté hors du noyau pour être ensuite traduit (s'il s'agit d'un ARNm), il subit une série de modifications qui concourent à la production à partir du transcrit primaire d'un ARN dit mature, apte à assurer les fonctions auxquelles il est destiné. Dans l'ordre où elles interviennent on distingue pour les ARNm : l'addition



d'une coiffe à l'extrémité 5', l'épissage des introns et l'addition d'une queue de polyadénylation à l'extrémité 3'.

- La découverte de la régulation des gènes de l'opéron lactose a mis en évidence pour la première fois, l'existence de séquences ADN régulatrices distinctes des séquences codantes.
- Les séquences ADN régulatrices sont reconnues par des protéines spécifiques activatrices ou répressives. Une même protéine régulatrice peut être à la fois répressive et activatrice selon les circonstances.
- Contrairement aux gènes bactériens dont les séquences sont continues, la plupart des gènes eucaryotes ont une séquence organisée en régions codantes (les exons) interrompues par des régions non codantes (les introns). Ils sont transcrits en une copie ARN continue, le transcrit primaire (appelé aussi pré-ARNm, s'il s'agit du précurseur de l'ARNm). Le pré-ARN est converti en ARN mature grâce à une machine moléculaire, le splicéosome.
- La régulation de l'épissage alternatif des ARN messagers a pour conséquence principale d'accroître la diversité des protéines cellulaires et d'aboutir à la synthèse de protéines différentes selon les types cellulaires considérés.
- L'hérédité épigénétique repose sur la transmission d'une information qui n'est pas codée dans la séquence de l'ADN bien qu'elle soit transmise de cellule-mère en cellule-fille et de génération en génération.
- Les deux principaux composants biochimiques du code épigénétique héréditaire sont les mécanismes de méthylation-déméthylation de l'ADN au niveau notamment des îlots CpG et les modifications des queues d'histones par des mécanismes d'acétylation-désacétylation et de méthylation-déméthylation.
- Chez les organismes diploïdes, les cellules somatiques possèdent deux copies de génome et chaque gène autosomal est donc représenté par deux copies (deux allèles), chacune héritée de l'un des deux parents au moment de la fécondation. Pour la plupart des gènes autosomaux, les deux allèles généralement s'expriment. Cependant, certains gènes (< 1 %) sont porteurs d'une empreinte parentale qui détermine lequel des deux allèles pourra s'exprimer.
- La génomique étudie la totalité du génome d'un organisme à la différence de la génétique moléculaire qui s'attache plus spécialement à l'étude de gènes particuliers en accordant une attention à leurs divers rôles et fonctions. L'une de ses démarches principales consiste à établir la séquence complète des génomes étudiés (génomique structurale) et à procéder à l'analyse de l'expression dans sa globalité (génomique fonctionnelle) en fonction notamment les conditions environnementales. Des approches similaires sont pratiquées également avec l'ensemble des transcrits (transcriptome) d'une cellule ou d'un organisme, c'est la transcriptomique, ou avec les protéines (protéome), c'est la protéomique.



- De nombreuses séquences génomiques sont conservées entre les espèces témoignant sans doute à la fois de l'avantage sélectif qu'elles représentent et de leur importance pour le fonctionnement du vivant. Mais toutes sont loin de coder des protéines, beaucoup sont transcrites en ARN non codants, ce qui a pour signification que le fonctionnement des organismes vivants semble étrangement dépendre dans une mesure aux limites encore inconnues de nombreux ARN dotés de fonctions régulatrices.
- Les organismes multicellulaires fonctionnent dès la première cellule-œuf et les premiers stades de développement, sur la base de réseaux de gènes qui assurent leur adaptation permanente aux conditions environnementales.

QCM - QROC

5.1 La séquence ci-dessous appartient à un segment d'ADN en cours de transcription. La flèche indique le sens de transcription de ce segment.

Brin A :

```
5'-CTAGTAAGGAGGTGCGAATGCTATGAAAGCAACTAAACTAGTACTTGGGGCGGTGAT-3'
 ———————>
```

Brin B :

```
3'-GATCATTCCTCCACGCTTACGATACTTTCGTTGATTTGATCATGAACCCCGCCACTA-5'
```

Quel est le brin matrice ?

Quel est le brin codant ?

Donner la séquence du transcrit.

5.2 L'ADN d'un gène morcelé est hybridé avec l'ARNm mature correspondant. Il se forme alors une structure avec des régions complémentaires (ARN-ADN) séparées par des boucles d'acide nucléique monocaténaire. Parmi les propositions suivantes, lesquelles sont bonnes ?

a) Les boucles sont faites d'ADN.

b) Les boucles correspondent aux exons du gène.

c) Quand ils sont hybridés, l'ADN et l'ARN sont antiparallèles.

d) Le brin d'ADN qui s'hybride à l'ARNm est le brin codant.

5.3 Considérons un gène eucaryote de 1 500 pb qui code pour une protéine de 300 amino-acides. Peut-on prédire les tailles respectives du transcrit primaire, des éventuels introns, de l'ARNm et des régions 3' et 5' non codantes ?

5.4 Une cellule est exposée à des agents mutagènes. Une protéine produite par cette cellule est examinée avant et après le traitement.



Il s'avère que dans la cellule traitée, la protéine contient un tryptophane (Trp) à la place d'une arginine (Arg).
En supposant que cette substitution soit due à une mutation ponctuelle, quelle est cette mutation ?

5.5 Compléter toutes les cases du tableau

<table>
<tr><td rowspan="2">ADN</td><td>3′</td><td>C</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>5′</td></tr>
<tr><td>5′</td><td></td><td></td><td></td><td></td><td></td><td></td><td>T</td><td>G</td><td>A</td><td></td><td></td><td></td><td>3′</td></tr>
<tr><td>ARNm</td><td>5′</td><td></td><td>C</td><td>A</td><td></td><td></td><td></td><td>U</td><td></td><td></td><td></td><td></td><td></td><td>3′</td></tr>
<tr><td>Anticodon de l'ARNt</td><td>3′</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>G</td><td>C</td><td>A</td><td>5′</td></tr>
<tr><td>Acide aminé incorporé dans la protéine</td><td>NH_2</td><td colspan="3"></td><td colspan="3">Trp</td><td colspan="3"></td><td colspan="3"></td><td>COOH</td></tr>
</table>

5.6 Quand une mutation par insertion ou délétion d'un ou deux nucléotides se produit dans un intron d'un gène, est-ce que cela affecte la séquence du transcrit et de la protéine qui pourrait en résulter ?

5.7 Valider les affirmations suivantes.

a) Les régions codantes et non codantes d'un gène contiennent des signaux de régulation de la transcription

b) Seules les régions codantes possèdent des signaux de régulation de la transcription

c) Seules les régions non codantes possèdent des signaux de régulation de la transcription

d) Les signaux de régulation peuvent se trouver hors des régions codantes et non codantes d'un gène

5.8 Comment expliquer que l'environnement influe sur l'expression des gènes ?

RÉPONSES

5.1 **a)** Brin B

b) Brin A

c)

`5′-CUAGUAAGGAGGUGCGAAUGCUAUGAAAGCAACUAAACUAGUACUUGGGGCGGUGAU-3′`

5.2 Les bonnes propositions sont : a et c. Les boucles correspondent aux introns et le brin d'ADN qui s'hybride à l'ARNm est le brin matrice.



5.3 La taille du transcrit primaire est de 1500 nucléotides. On ignore la présence et donc les tailles des introns. L'ARN messager comporte, après un éventuel épissage, 903 nucléotides (900 correspondent aux 300 amino-acides et 3 au codon d'arrêt). Les tailles des régions non codantes comportent au total 597 nucléotides,, introns éventuels compris. La taille des régions non codantes en 3' et 5' ne peuvent pas être déterminées sur la base des informations fournies.

5.4 Arg a les codons suivants : CGU, CGC, CGA, <u>CGG</u>, AGA, <u>AGG</u>
Trp a un seul codon, <u>UGG</u>.

Aucun codon de Arg ne commence par U. La mutation doit par conséquent concerner cette position du codon. Étant donnée que la mutation est ponctuelle, il ne peut donc s'agir que du changement C → U ou A → U. Au niveau de l'ADN c'est donc la paire C-G qui est mutée en paire T-A ou la paire A-T qui devient T-A.

5.5

ADN	3′	**C**	G	T	A	C	C	A	C	T	G	C	A	5′
	5′	G	C	A	T	G	G	**T**	**G**	**A**	C	G	T	3′
ARNm	5′	G	**C**	**A**	U	G	G	**U**	G	A	C	G	U	3′
Anticodon de l'ARNt	3′	C	G	U	A	C	C	*	*	*	**G**	**C**	**A**	5′
Acide aminé incorporé dans la protéine	NH_2	Ala			**Trp**			*			-			COOH

5.6 L'insertion ou délétion d'un ou deux nucléotides dans un intron d'un gène peut ne rien modifier dans les séquences de l'ARNm et de la protéine. Cependant si cette mutation modifie les sites donneurs et accepteurs d'épissage, ou le point de branchement d'un intron, l'épissage en sera affecté et la séquence intronique pourra se retrouver dans l'ARNm. Dans ce cas, les séquences de l'ARNm et de la protéine qui en résulte seront modifiées.

5.7 **a)** et **d)**

5.8 En exerçant par divers mécanismes une modulation du niveau et/ou de l'activité des protéines régulatrices (activateurs et répresseurs) de la transcription.



Transmission et hérédité

PLAN

OBJECTIFS

- Connaître les mécanismes de recombinaison génétique qui génèrent la diversité des gamètes
- Comprendre la liaison génétique et savoir cartographier des gènes
- Savoir identifier la transmission de caractère liée au chromosome sexuel
- Comprendre en quoi les interactions génétiques modifient la transmission des caractères

6.1 LES DIVISIONS CELLULAIRES

La mitose et la méiose constituent les deux modes de division chez les eucaryotes.

La mitose

La mitose concerne à la fois les cellules haploïdes et diploïdes. Elle aboutit à deux cellules filles génétiquement identiques à la cellule mère *(Fig. 6.1)*, le nombre de chromosome par cellule demeurant constant. La mitose s'effectue après la réplication de l'ADN qui s'accomplit au cours de la phase S du cycle cellulaire (voir dans la même collection, *Biologie cellulaire*). Elle se subdivise en quatre phases :

- la **prophase** initie la mitose avec le début de la condensation de la chromatine, et l'individualisation des chromosomes. À ce stade, les

deux chromatides sœurs sont accolées sur toute leur longueur, et plus particulièrement au niveau du **centromère**, région d'hypercondensation de la chromatine, auxquels viennent se fixer de part et d'autre des protéines formant des complexes protéiques, appelés **kinétochores** ;

➤ la **métaphase** ou alignement des chromosomes est un point clé de la mitose. Tous les chromosomes ont leurs deux kinétochores liés aux microtubules, et s'alignent dans le plan équatorial de la cellule. L'équilibre établit entre les deux chromatides d'un même chromosome les maintient et les oriente vers les pôles opposés de la cellule ;

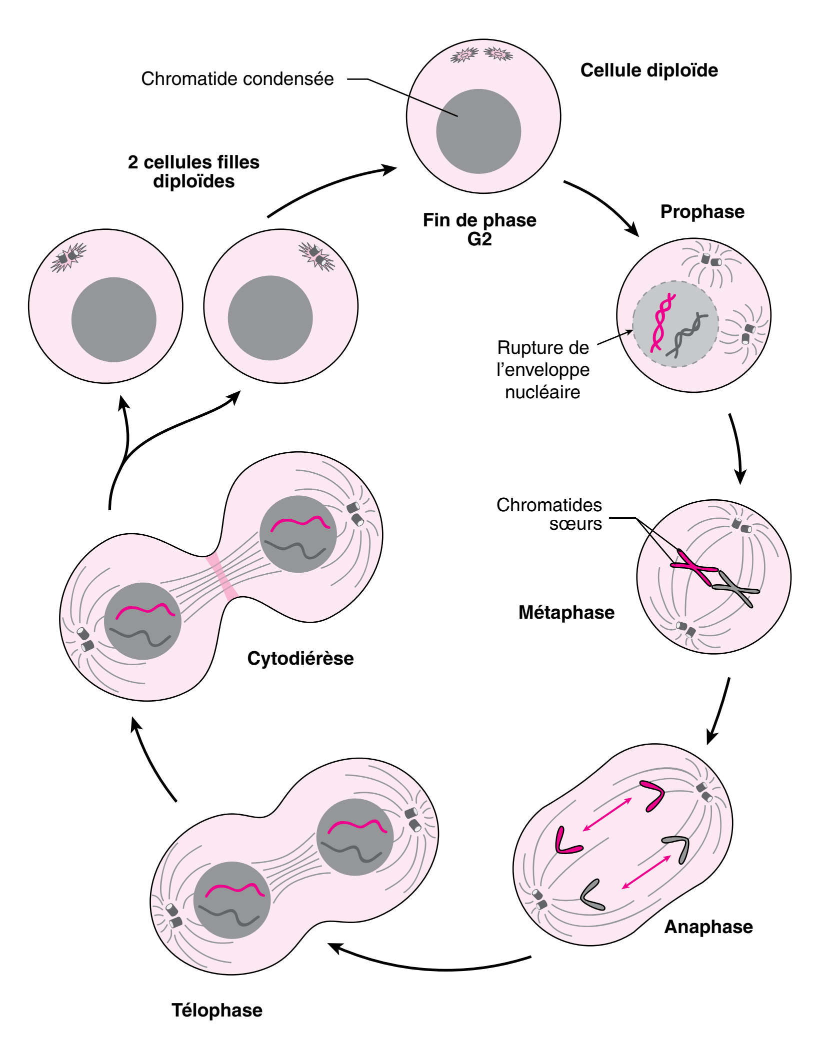


Figure 6.1 Représentation schématique des différentes phases de la mitose.



- l'**anaphase** assez rapide ne dure que quelques minutes pendant lesquelles les chromatides, devenues maintenant des chromosomes, migrent vers les pôles opposés de la cellule ;
- la **télophase** se caractérise par le regroupement des chromosomes fils aux pôles de la cellule qui marque la fin de la mitose. Des enveloppes nucléaires se reforment autour de chaque lot de chromosomes fils qui commencent à se décondenser.

Ainsi, les deux cellules filles héritent d'un nombre égal de chromosomes (n pour les cellules haploïdes et 2 *n* pour les cellules diploïdes).

La méiose

La **méiose** ne peut exister que pour des cellules obligatoirement diploïdes et aboutit à partir d'une cellule mère diploïde à quatre cellules haploïdes génétiquement différentes entre elles *(Fig. 6.2)*. La réduction du contenu en chromosomes résulte de l'existence d'une seule réplication de l'ADN suivie par deux cycles de divisions méiotiques :

- la première division méiotique ou **méiose I**, au cours de laquelle les chromosomes dont les chromatides restent associées ségrègent à chacun des pôles de la cellule lors de l'anaphase I. Il y a réduction du nombre de chromosomes, leur nombre étant divisé par 2, à la suite de cette première division. On parle alors de **division réductionnelle** ;
- la deuxième division méiotique ou **méiose II** durant laquelle les chromatides sœurs se séparent et migrent aux pôles de la cellule selon un mode égal de répartition. On parle de **division équationnelle**.

Lors de la première division méiotique les chromosomes homologues s'apparient formant des structures appelées **bivalents**. À ce stade les chromatides sœurs dupliquées restent étroitement associées, et s'apparient avec leurs homologues pour former des structures à 4 chromatides. L'appariement est un processus complexe qui permet à chaque cellule fille d'hériter d'un lot haploïde de chromosomes comprenant un seul exemplaire de chaque paire d'homologues. Il se déroule au cours de la prophase I, et comporte :

- Le **stade leptotène**, marqué par le début de l'individualisation et de la condensation des chromosomes. Les chromatides sœurs sont associées entre elles par l'intermédiaire de protéines cohésives et formeront les structures axiales du **complexe synaptonémal**. Des appariements transitoires favorisent l'alignement des homologues entre eux. Les événements de recombinaison homologue sont initiés et se caractérisent par des cassures des brins de la molécule d'ADN qui permettent des interactions entre les brins d'ADN appartenant à des chromatides homologues. En certains points de cassures se produisent des interactions conduisant aux **crossing-over** et constituent

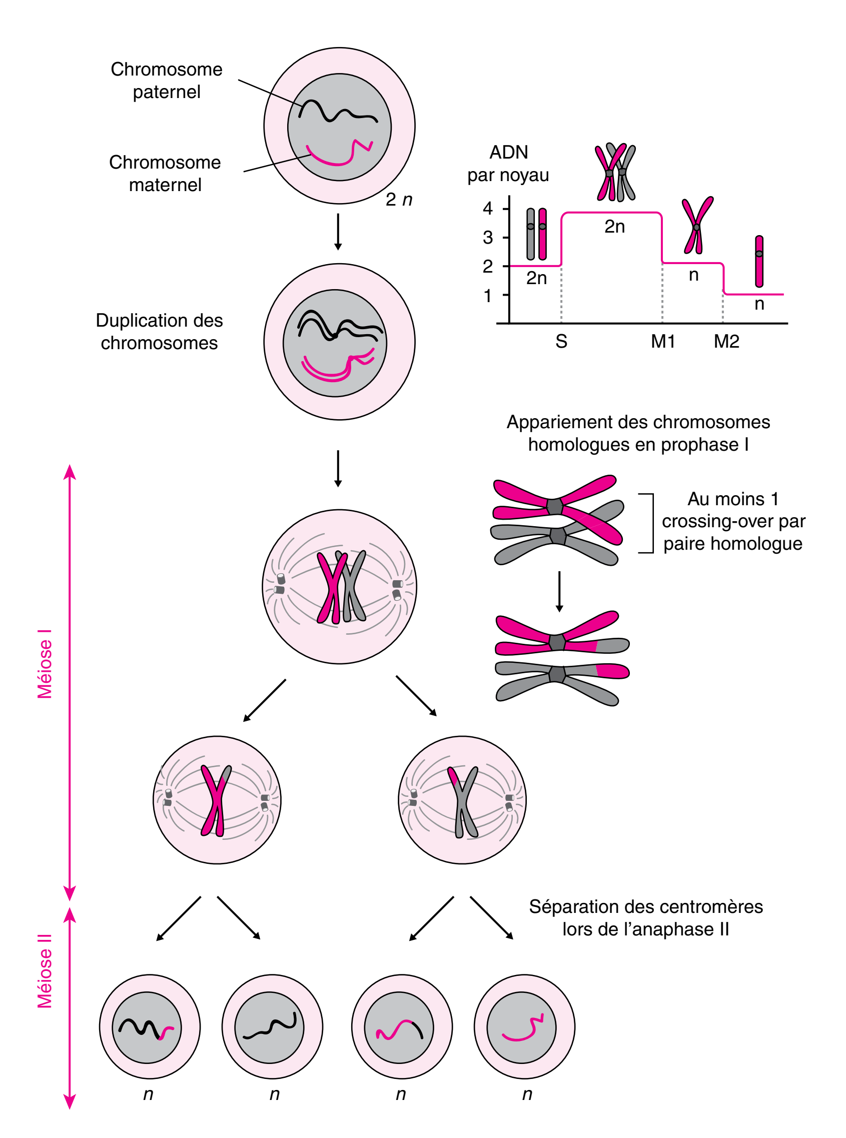


Figure 6.2 Représentation schématique de la méiose.

Seules sont représentées les division réductionnelles de la 1re division méiotique et équationnelle de la 2e division. Le graphique montre les variations du contenu en ADN et en chromosome au cours de la méiose. La cellule diploïde contenant 2 *n* chromosomes donne quatre cellules haploïdes à n chromosomes. Pour plus de clarté, seule la ségrégation d'un chromosome est montrée ainsi que celle de ses chromatides lors de la 2e division.



des **chiasmas**. Le crossing-over nécessite une cassure des doubles hélices d'ADN maternelles et paternelles de chacune des deux chromatides impliquées et leur réunion. Ces premiers points d'ancrage servent ultérieurement pour la formation des éléments centraux des complexes synaptonémaux constitués de protéines qui, sur toute la longueur du bivalent, maintiendront les homologues appariés.

➤ Les **stades zygotène** puis **pachytène** correspondent respectivement à la propagation de l'élément axial du complexe synaptonémal et l'appariement *via* l'élément central des chromosomes homologues sur toute leur longueur.

➤ Le stade **diplotène** coïncide avec la disparition du complexe synaptonémal, et la perte de l'alignement étroit des chromosomes homologues qui restent connectés entre eux au niveau des chiasmas.

La séparation des homologues à l'anaphase I ne pourra s'effectuer qu'après disparition des connexions assurées par les chiasmas, seule celle au niveau du centromère sera maintenue empêchant la séparation des chromatides d'un même chromosome. Par la suite, la méiose II s'effectuera après une courte interphase sans aucune réplication préalable de l'ADN.

La recombinaison méiotique

La **recombinaison méiotique** désigne tout processus méiotique qui donne un produit haploïde présentant une combinaison d'allèles différente de celle portée par les génotypes haploïdes à l'origine du **méiocyte**, cellule dans laquelle s'effectue la méiose, cellules de la lignée germinale chez les mammifères. Cette recombinaison est détectée en comparant les génotypes des descendants (génotype de « sortie ») par rapports à ceux des parents (génotypes d'« entrée »). Ainsi, tout produit méiotique avec une nouvelle combinaison des allèles fournis par les deux génotypes parentaux est par définition un recombinant.

Il s'avère facile de détecter les recombinants chez les organismes dont les cycles vitaux sont haploïdes comme la levure (voir *Fig. 4.12, chapitre 4*) car les génotypes peuvent être déduits directement des phénotypes. En revanche, pour les organismes diploïdes, détecter les recombinants est beaucoup plus compliqué. On se doit de connaître les génotypes des gamètes de départ et des gamètes résultants pour avoir accès aux recombinants. La démarche consiste à obtenir à partir de lignées pures (parents P1) pour les caractères concernés une génération F1 hétérozygote (par exemple de génotype *AaBb*), puis de croiser les individus F1 avec un individu double récessif (de génotype *aabb*) ayant une seule possibilité de génotype pour ses gamètes (*ab*), ce parent est également appelé **testeur** *(Fig. 6.3)*. L'analyse de la descendance à l'issue d'un tel croisement appelé croisement test ou

« test-cross » permet d'identifier les recombinants dont les phénotypes diffèrent de ceux des parents P1. Les recombinants sont produits par deux processus distincts : l'assortiment indépendant et les crossing-over.

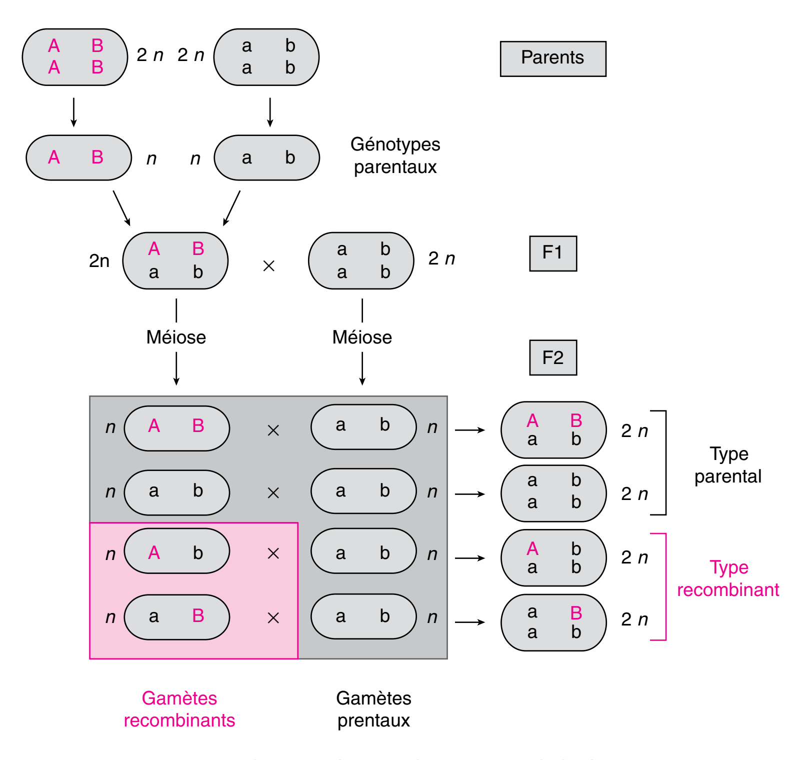


Figure 6.3 Détection des recombinants d'une méiose diploïde par un croisement test.

Les individus de la F1 ayant tous le même génotype, leur croisement avec un double récessif permet d'obtenir à la génération F2 des individus dont le génotype peut se déduire du phénotype et de les classer selon leur type parental ou recombinant.

La recombinaison interchromosomique

Le brassage interchromosomique correspond à une redistribution aléatoire des chromosomes parentaux lors de l'anaphase de la première division méiotique. Ainsi dans l'exemple représenté dans la *figure 6.4*, on considère 2 paires de chromosomes homologues pour chaque parent. On constate pour certains gamètes recombinants de la génération F1, un tel brassage interchromosomique. Ce processus engendre une grande diversité de gamètes, car il existe 2^n possibilités de répartition de chromosomes, soit chez l'homme 2^{23} ($8,4 \times 10^6$) gamètes susceptibles d'être génétiquement différents.

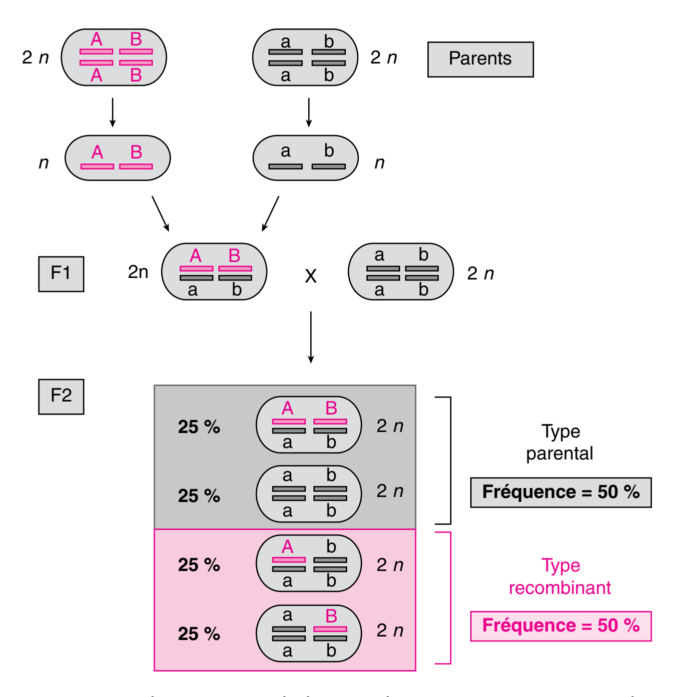


Figure 6.4 Recombinants issus de la recombinaison entre gènes non liés à la suite d'un assortiment indépendant.

Chaque paire d'homologue porte un des caractères étudiés.

Lorsque les gènes sont situés sur des chromosomes différents, les gamètes sont issus d'un assortiment indépendant. Les génotypes « *Aabb* » et « aaBB » sont de type recombinant car ils diffèrent des génotypes parentaux de départ « *AABB* » et « *aabb* ». La fréquence des recombinants à la suite d'un assortiment indépendant doit être de 50 %. Une telle valeur obtenue après un croisement-test laisse supposer que l'assortiment des gènes étudiés est indépendant, c'est-à-dire que les gènes sont portés par des paires de chromosomes distincts.

La recombinaison intrachromosomique

En second lieu, la diversité des gamètes peut être accrue en raison des enjambements chromosomiques ou « crossing-over » *(Fig. 6.2* et *6.5)*. C'est la **recombinaison intrachromosomique**, elle accroît encore un peu plus la diversité génétique des populations ultérieures.

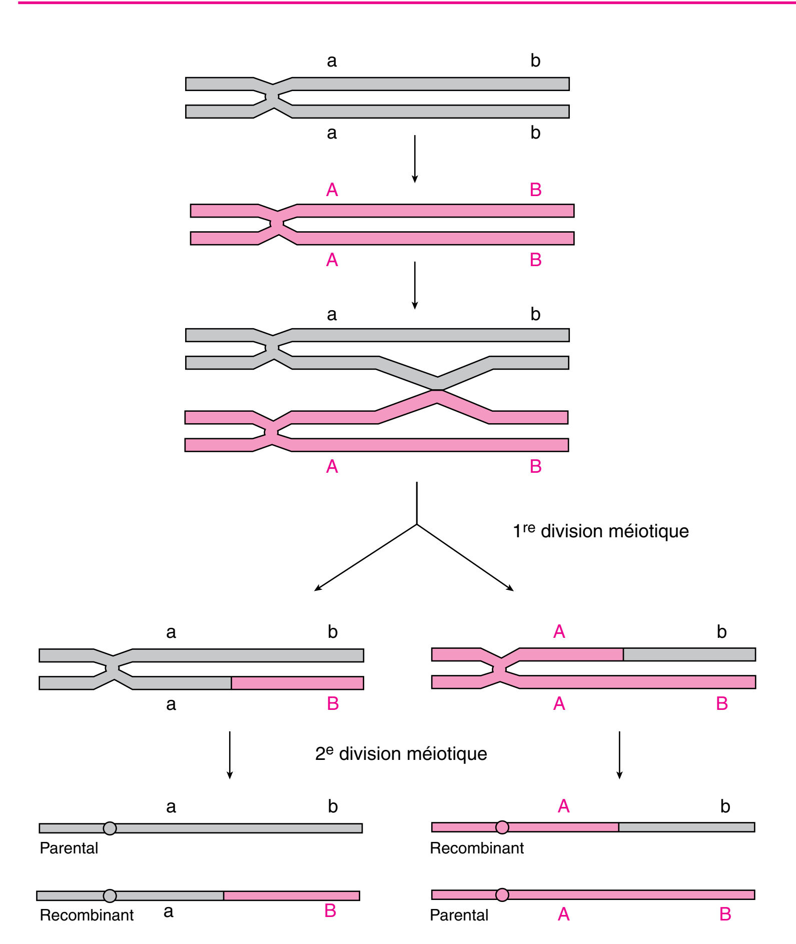


Figure 6.5 Crossing-over au cours de la méiose et production de gamètes recombinants.

Dès le début du XXe siècle, les travaux sur l'hérédité menés par William Bateson, avec comme modèle le pois de senteur, et ceux de Thomas Hunt Morgan, avec comme modèle la drosophile, avaient montré que la ségrégation des caractères n'obéissait pas toujours aux lois de Gregor Mendel.

Prenons comme exemple un croisement-test où les individus F1 sont croisés avec le parent double récessif, les résultats obtenus sont présentés dans le *tableau 6.1*.



TABLEAU 6.1 EXEMPLE DE L'EFFECTIF DES DESCENDANTS À L'ISSUE D'UN CROISEMENT-TEST POUR DEUX GÈNES LIÉS.

Combinaisons alléliques gamètes possibles		Caractéristiques des individus de la génération F2		Nombre d'individus	
F1	**P2**	**Phénotypiques**	**Génotypiques**	**Observés**	**Attendus sur la base du rapport 1:1:1:1**
AB	ab	AB	AaBb	1 402	750
Ab	ab	Ab	Aabb	98	750
aB	ab	aB	aaBb	102	750
ab	ab	ab	aabb	1 398	750

Les différentes combinaisons alléliques des gamètes de la F1 sont les suivantes : *AB*, *Ab*, *aB* et *ab*. Par contre, le testeur ne transmet que les deux allèles récessifs, c'est-à-dire une seule combinaison allélique (*ab*). Aussi, les différents phénotypes de la génération F2, obtenus par le croisement F1 × P2 (parent double homozygote récessif) dépendront uniquement de la contribution gamétique de F1. Les individus sont comptés et classés selon leur phénotype (ou génotype déduit). On peut remarquer que :

- Les valeurs observées pour chaque classe phénotypique diffèrent largement du rapport mendélien 1:1:1:1 attendu, si la ségrégation des caractères avait été indépendante.
- Deux classes phénotypiques prédominent, l'une correspond aux individus porteurs de la combinaison des deux caractères dominants (*A* et *B*) et l'autre à ceux possédant les caractères récessifs (*a* et *b*). Le nombre d'individus de ces deux sous-populations représente 93,4 % de la population totale, le rapport entre les deux sous-populations : individus de phénotype *AB* (1 402) et individus de phénotype *ab* (1 398) étant voisin de 1. Il faut remarquer que ces phénotypes F2 correspondent aux deux combinaisons gamétiques (*AB* et *ab*) provenant des parents et préalablement transmis à la génération F1.
- Deux classes d'individus ayant les phénotypes respectifs *Ab* et *aB* sont minoritaires et ne représentent que (98 + 102)/3 000 = 6,6 % de la population F2. Là encore, on notera que le rapport entre les deux sous-populations (98/102) est voisin de 1.

Pour expliquer ces résultats, les généticiens emploient l'expression « couplage physique » entre les allèles dominants (ici *A* et *B*) d'une part et les allèles récessifs (*a* et *b*) d'autre part. Cette liaison physique empêche que leur ségrégation soit indépendante à l'exception d'une très faible proportion. Thomas Morgan proposa que les gènes étudiés se localisent sur une paire de chromosomes homologues *(Fig. 6.6)*, l'un portant une combinaison allélique, par exemple les allèles dominants *A* et *B*, et son homologue la combinaison allélique associant les deux allèles récessifs *a* et *b*. On parle pour qualifier cette situation de **liaison génétique**. Les événements de recombinaison intrachromosomiques qui peuvent se produire entre la région portant les deux caractères étudiés étant à l'origine des quelques recombinants observés.

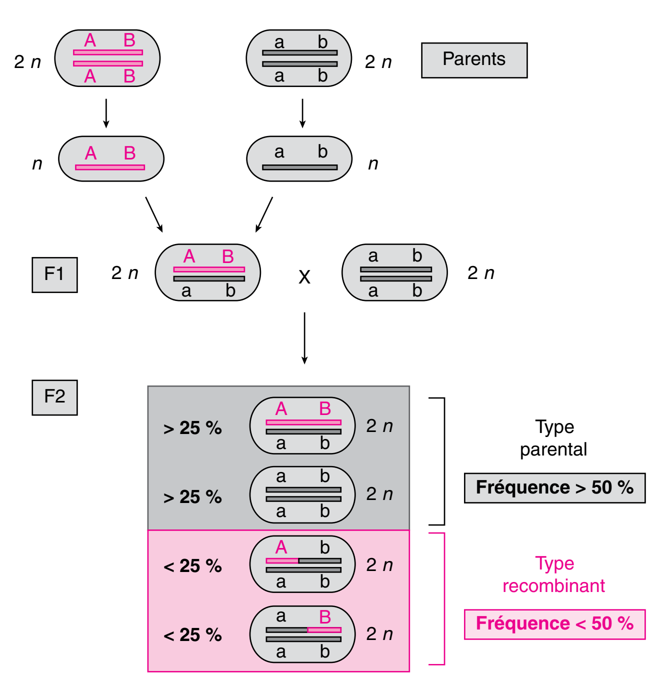


Figure 6.6 Recombinaison entre gènes liés et fréquence des gamètes. Un seul chromosome porte les 2 caractères étudiés.

La recombinaison intrachromosomique peut se produire entre n'importe lesquelles des deux chromatides non sœurs *(Fig. 6.7)*. Cependant le pourcentage de recombinants est toujours inférieur à 50 % en raison de leur rattachement au même chromosome, ce qui empêche leur assortiment indépendant. On doit de plus noter que la valeur seuil de 50 % n'implique pas forcément l'assortiment indépendant des



caractères. En effet, des gènes liés mais très distants sur un même chromosome auront cependant une fréquence maximale de recombinaison de 50 %. Pour l'observateur tout se passe comme si des gènes disposés ainsi étaient indépendants, c'est-à-dire sur deux chromosomes différents. C'est la forte probabilité qu'il y ait au moins un crossing-over dans le grand intervalle qui sépare les deux gènes qui favorise une transmission indépendante de ces gènes, bien que physiquement ils soient sur le même chromosome.

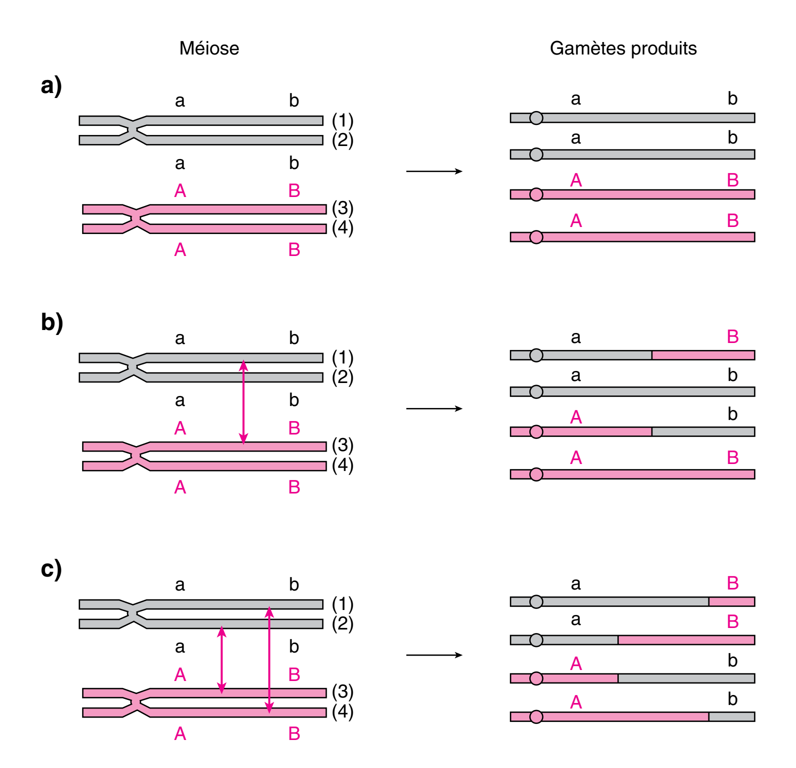


Figure 6.7 Exemples de possibilités de crossing-over.

Les gamètes produits sont tous de type parental en absence de crossing-over **(a)** ; de type parental et recombinant lors de crossing-over impliquant les chromatides 1 et 3 **(b)**, ou 1 et 4, et 2 et 3 **(c)**. Ce dernier exemple montre que lors de la méiose les crossing-over peuvent s'effectuer entre n'importe lesquelles des chromatides non sœurs et que plusieurs crossing-over peuvent s'effectuer pour un même bivalent.

6.2 LIAISON GÉNÉTIQUE ET CARTOGRAPHIE

Chaque molécule d'ADN composant un chromosome, contient des séquences informatives ou gènes disposés linéairement. Aussi, la cartographie d'un génome consiste à :



- localiser ou assigner les gènes sur les chromosomes et établir les distances qui séparent les gènes, c'est la cartographie génétique :
- reconstituer le génome à l'échelle moléculaire en établissant au final la totalité de la séquence, c'est la cartographie physique.

Naturellement, les cartographies physique et génétique, tout comme le séquençage sont interdépendants et leur association ont favorisé l'avancée rapide de la connaissance des génomes complexes comme celui de l'homme.

Fréquence de recombinaison intrachromosomique et distance génétique

En augmentant le nombre de caractères (ou gènes) étudiés, on constate que le pourcentage de recombinants parmi la descendance peut varier de 0 à 50 % selon la localisation du couple de caractères observés. Comme la fréquence de recombinaison entre deux loci reflète la distance génétique qui les sépare, on peut évaluer ainsi les distances entre les gènes qui déterminent les caractères : plus le nombre de recombinants est élevé, plus il y a en théorie de chance pour que les gènes soient éloignés sur le même chromosome. En résumé, évaluer les distances entre les gènes sur un même chromosome revient à calculer le nombre de recombinants pour différents couples de caractères dans la descendance. C'est pourquoi on utilise le pourcentage de recombinants comme une mesure de la distance linéaire séparant deux gènes. On peut ainsi établir une carte génétique linéaire qui donne la position relative des différents loci ou gènes les uns par rapport aux autres. Sur la carte, les distances sont mesurées en unité de carte génétique ou UG (Unité Génétique) : une UG étant la distance séparant deux loci dont la fréquence de recombinaison est de 1 % (c'est-à-dire 1 produit sur 100 de la méiose est un recombinant). L'unité la plus couramment employée est le Morgan (en l'honneur de Thomas Hunt Morgan). On parle en général de centimorgan (cM) ; 1 cM équivaut à une fréquence de recombinaison de 1 %, soit 1 UG (chez l'homme, on considère qu'une distance d'un cM équivaut en moyenne à une séquence d'ADN de 10^6 pb). Chez le ver nématode, 1 cM équivaut à $7{,}5 \times 10^5$ pb.

Comment cartographier plus de deux gènes liés

Quand on examine le nombre de recombinants pour différents couples de caractères, on établit des distances génétiques entre les loci composant ces couples. Cependant, si un locus *C* est commun à deux couples (*AC* et *BC*), comment établir la cartographie de la région *ABC* ? En effet, si *A* et *C* sont distants de 7 cM, et *B* et *C* de 10 cM, plusieurs combinaisons s'avèrent possibles, comme présenté dans la *figure 6.8*.



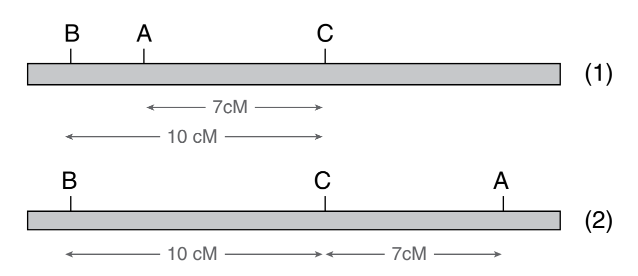


Figure 6.8 Position envisagée pour différents loci portés par un même chromosome. Les distances entre les caractères sont données en centimorgans (cM).

Dans le cas (1) on postulera une distance entre *A* et *B* de 10 – 7 = 3 cM, alors que dans le cas (2) *A* et *B* seront distants de 10 + 7 = 17 cM. Avec le gène *B* comme point de départ, deux cartes sont possibles respectivement *BAC* (1) ou *BCA* (2). Pour lever l'ambiguïté, il s'avère nécessaire d'analyser les couples deux à deux, soit dans notre exemple les couples de loci *AB*, *BC*, *AC*. Considérons que *ABC* sont des allèles dominants et *abc* des allèles récessifs. Les parents P_1 et P_2 donnant la F1 (de génotype *AaBbCc*) peuvent posséder, parmi d'autres, les génotypes respectifs suivants : *aaBBCC* et *AAbbcc*, auxquels on se référera par la suite.

Lors d'un croisement test de F1 avec un triple récessif *aabbcc*, les huit types gamétiques F1 haploïdes possibles peuvent être dénombrés grâce aux phénotypes des individus diploïdes de la génération F2, puisque le testeur gamétique n'apporte que le type gamétique *abc (Tableau 6.2)*.

Tableau 6.2 Classification et dénombrement en fonction de leurs phénotypes d'individus issus d'un croisement-test.

Phénotypes	Génotypes	Nombre d'individus
aBC	*aaBbCc*	1 250
Abc	*Aabbcc*	1 258
abC	*aabbCc*	143
ABc	*AaBbcc*	139
ABC	*AaBbCc*	93
abc	*aabbcc*	99
aBc	*aaBbcc*	8
AbC	*AabbCc*	10



Les phénotypes aBC et Abc correspondent aux types parentaux de départ. Ils sont à la génération F2 les populations les plus représentées (1 250 et 1 258 individus respectivement). Les autres populations ne comportent qu'un faible nombre d'individus qui résulte de recombinaisons (brassages intrachromosomiques) lors de la méiose aboutissant à la formation des gamètes de la génération F1.

La comparaison entre types parentaux et types recombinants permet d'identifier les événements de recombinaison entre deux loci. Par exemple pour les loci *A* et *B*, les recombinants qui diffèrent des types parentaux *aB* et *Ab* sont dans l'ordre *ab* et *AB* ; leur représentation dans les différentes populations recombinantes s'établit à 143 + 139 + 93 + 99 = 474 individus, soit une fréquence de recombinaison de 15,8 % (× 100). Pour les loci *A* et *C*, les recombinants sont *ac* et *AC,* avec au total 93 + 99 + 8 + 10 = 210 individus soit une fréquence de 7 %. Enfin pour *B* et *C*, les recombinants sont *Bc* et *bC*, avec au total 143 + 139 + 8 + 10 = 300 individus soit une fréquence de 10 %. Dans les trois cas les fréquences de recombinaison sont inférieures à 50 %, les trois gènes sont liés et situés sur le même chromosome, les deux gènes les plus distants étant *A* et *B* avec 15,8 % de recombinaison soit 15,8 cM. À ce stade, nous pouvons préciser la carte génétique pour ces trois gènes.

L'ordre des gènes défini arbitrairement en (2) en plaçant *A* à droite de *C*, correspond à celui déterminé par l'analyse des fréquences de recombinaison. On doit remarquer que pour cet ordre la somme des distances entre *AC* et *BC* (7 + 10 = 17 cM) diffère malgré tout de celle calculée à partir des recombinaisons entre les loci *A* et *B* (15,8 cM). Pour expliquer cette différence, il faut prendre en considération les deux sous-populations les moins représentées ayant les phénotypes : aBc et AbC. Pour les caractères *A* et *B*, nous retrouvons les deux combinaisons parentales *aB* et *Ab*. Dans la *figure 6.9*, on décrit la possibilité que deux « crossing-over » se soient produits, l'un entre *A* et *C*, et l'autre entre *C* et *B* conduisant à des doubles recombinants. Leurs phénotypes pour les caractères *A* et *B* étant identiques à ceux observés chez les parents, nous avons exclu ces deux sous-populations rares du comptage des recombinants formés entre *A* et *B*. Puisqu'il s'agit de doubles recombinants, nous devons au contraire les comptabiliser deux fois. Ainsi la fréquence de recombinaison (FR) entre les caractères *A* et *B* devient :

$$FR = [143 + 139 + 93 + 99 + 2\,(8 + 10)] \times 100 = 17\ \%$$

La distance séparant *A* et *B* s'élève maintenant à 17 cM, ce qui correspond bien à la somme des deux plus petites distances calculées ci-dessus.

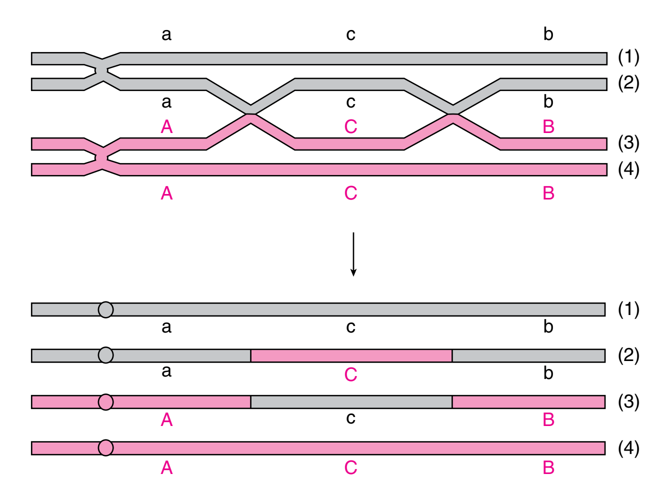


Figure 6.9 Exemple de double « crossing-over ».

Dans l'exemple choisi ici, nous avons pu mettre en évidence des doubles recombinants, car chaque « crossing-over » est entouré par une paire de gènes possédant deux allèles distincts pour chaque caractère. Si le caractère *C* avait été homozygote (allèles *C*), les doubles recombinants auraient eu comme phénotypes aBC et AbC, c'est-à-dire identiques pour les trois caractères aux types gamétiques parentaux de départ : (*aaBBCC* et *AAbbCC*). Ils auraient alors été classés par erreur dans les types parentaux lors de l'identification et du comptage des individus F2.

Remarque : lorsque la distance entre deux loci est relativement grande, l'existence de deux crossing-over, voire plus, peut entraîner une sous-estimation de la distance réelle les séparant. Il s'avère donc nécessaire d'avoir de nombreux caractères (marqueurs) répartis à des intervalles réguliers sur le chromosome, afin d'obtenir une carte génétique la plus proche possible de la réalité.

Les sous-populations de doubles recombinants ont des effectifs très faibles comparés aux autres populations. C'est pourquoi un seul arrangement des gènes sur le chromosome, parmi les différents possibles, peut donner ces sous-populations rares lors de double crossing-over. Ce qui permet, en examinant la composition de la descendance, d'en déduire directement l'ordre des trois gènes en confrontant les différentes combinaisons théoriques possibles *(Fig. 6.10)*. On voit que, parmi les trois arrangements théoriques, seul l'ordre (3) donne des doubles recombinants compatibles avec les plus faibles effectifs observés dans les sous-populations de la descendance soit *acB* (8 individus) et *ACb*

(10 individus). Ce résultat est en accord avec le dénombrement des différentes sous-populations et avec l'hypothèse (2) de départ.

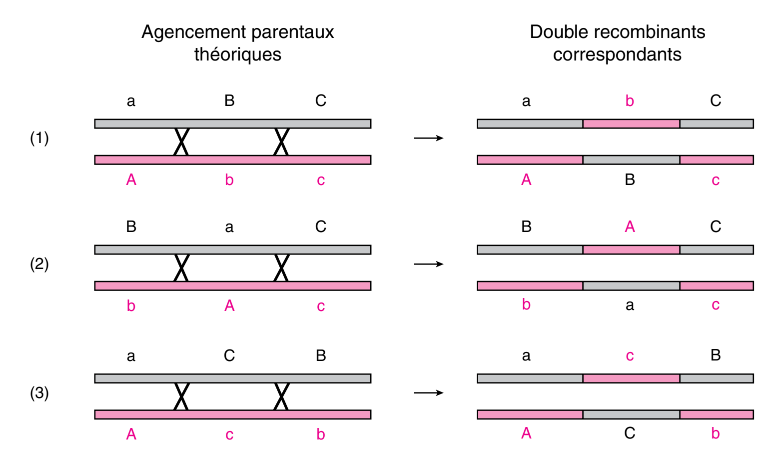


Figure 6.10 Différentes combinaison de doubles recombinants.

Pour retrouver les génotypes parentaux (*aB* ou *Ab*), seule la combinaison 3 permet de l'obtenir avec deux crossing-over, chacun d'eux s'effectuant entre des couples de caractères différents.

La notion d'interférence entre crossing-over

À partir des fréquences de recombinaisons établies entre les différentes paires de gènes, il devient possible d'estimer, de manière théorique, pour la région concernée, le pourcentage de doubles recombinants. Si la fréquence de recombinaison entre *A* et *B* est de 7 % et celle entre *A* et *C* de 10 %, la fréquence d'apparition des doubles recombinants devrait être égale à $0{,}07 \times 0{,}10 = 0{,}007 = 0{,}7$ %. Ce qui correspond à un nombre théorique d'individus de $(0{,}7 \times 3\,000)/100 = 21$ individus doubles recombinants. Or dans notre exemple, sur les 3 000 individus composant la génération F2, on en dénombre 18. La différence suggère que, dans les régions proches, il n'y a pas d'indépendance dans la réalisation du crossing-over. Il existe une interférence (I) qui s'explique par le fait que l'existence d'un crossing-over diminue la probabilité qu'un deuxième crossing-over se produise dans une région très proche. I peut s'exprimer ainsi :

$$I = 1 - \frac{\text{nombre de doubles recombinants observés}}{\text{nombre de doubles recombinants attendus}}$$

Le rapport entre les nombre de recombinants observés et attendus est appelé coefficient de coïncidence (cdc), donc I = 1 – cdc.

Quand I = 1, l'interférence est maximale, il n'y a pas de doubles recombinants observés dans la région concernée. En revanche si I = 0, le nombre attendu est identique au nombre théorique, donc le rapport ou coefficient de coïncidence est égal à 1. En général, la valeur de I se situe entre 0 et 1, dans notre exemple I = 1 – 18/21, I = 1 – 0,85 d'où I = 0,15. Cette valeur de l'interférence reflète l'écart entre le nombre d'individus observés et le nombre d'individus attendus. L'existence d'interférences désigne une plus grande fréquence des simples crossing-over dans une région chromosomique.

6.3 CARTOGRAPHIE DES CENTROMÈRES ET ANALYSE DES TÉTRADES LINÉAIRES

Les centromères, régions des chromosomes qui ne présentent pas d'hétérozygotie, sont difficiles à cartographier. Cependant, chez des champignons comme *Neurospora crassa* dont les produits de la méiose sont ordonnés dans l'asque, les centromères peuvent être cartographiés. En effet, le méiocyte produit une tétrade linéaire, puis chaque spore subit une mitose postméiotique *(Fig. 6.11)*. Les huit ascospores, constituent une **octade**, et sont réparties dans l'asque en fonction du profil de ségrégation résultant des deux divisions méiotiques. Ainsi, on peut utiliser la cartographie du centromère pour estimer la distance qui sépare un locus du centromère d'un chromosome. Cette technique est basée sur le fait qu'à la suite d'un crossing-over lors d'une méiose

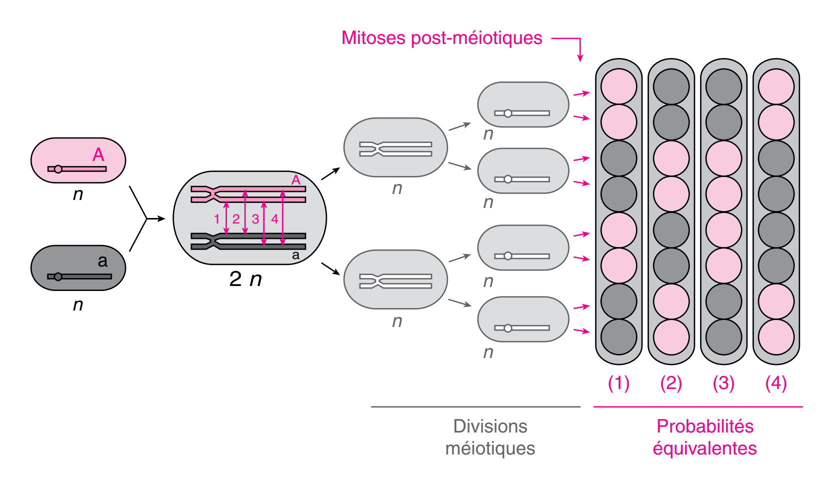


Figure 6.11 Les quatre différents profil de ségrégation de seconde division.

À la suite d'un crossing-over impliquant deux chromatides non sœurs, les différentes combinaisons sont indiquées par les flèches, *A* et *a* ségrègent dans les noyaux distincts lors de la seconde division méiotique. La répartition des ascospores dans l'asque linéaire dépend des chromatides impliquées dans le crossing-over.



l'ordonnancement imposé aux ascospores sera modifié. Par exemple, en croisant deux individus ayant chacun un allèle différent (*a* ou *A*) à un locus et en absence de crossing-over dans la région comprise entre le locus et le centromère, on observera une octade comportant deux blocs adjacents de quatre ascospores (l'un de génotype *a* et l'autre de génotype *A*). En revanche, si un crossing-over s'effectue dans cette région, les ascospores seront réparties par bloc de deux en fonction des chromatides impliquées dans la recombinaison. On obtient ainsi quatre différents profils de répartition *(Fig. 6.11)*. On les appelle **profils de ségrégation de seconde division** (M_{II}) car les allèles *a* et *A*, du fait du crossing-over, se retrouvent dans les noyaux à la fin de la première division. Sans crossing-over, les noyaux à l'issue de la première division porteront les mêmes allèles (*aa* ou *AA*), on parle alors de **profils de ségrégation de première division** (M_I).

La fréquence des octades ayant un profil M_{II} doit être proportionnelle à la distance qui sépare le locus du centromère, elle peut donc servir pour estimer la taille de la région. Après un croisement entre individus *A* et *a*, on dénombre les octades de profil M_I et celles de profil M_{II}. Il en résulte que le pourcentage de recombinants est égal à :

$$\text{Nombre de MII/Nombre total d'ascospores } (M_I + M_{II})$$

En réalité, la valeur trouvée représente un nombre de méiose et non le nombre de chromatides recombinantes. Comme dans les méioses, seulement 50 % des chromatides sont recombinantes, on doit diviser la fréquence des M_{II} par 2 pour la convertir en unités génétiques.

Inversion chromosomique et cartographie

Chez la drosophile la **cytogénétique** permet de repérer les inversions chromosomiques. En effet, dans les glandes salivaires de ces insectes, on trouve des **chromosomes polytènes**, chromosomes géants à la suite de multiples duplications de leurs chromatides qui restent soudées donnant un profil de bandes caractéristiques de chaque chromosome. Les mutations responsables de réarrangements chromosomiques comme des inversions peuvent alors être facilement mises en évidence sur un tel chromosome polytène grâce à une modification de l'agencement des bandes *(Fig. 6.12)*. Si à cette mutation est associé un nouveau phénotype, le généticien peut déterminer la position approximative de l'allèle muté responsable de ce nouveau phénotype.

Grâce aux chromosomes polytènes, il a été possible de caractériser un gène homéotique, le gène *bithorax* (*BX-C*) par une marche sur le chromosome. Les auteurs avaient remarqué que des inversions dans la région 89E du chromosome 3 affectaient profondément l'embryon de drosophile. Cette région devait correspondre à un gène majeur du développement. Cependant, ils ne disposaient pas de sonde pour

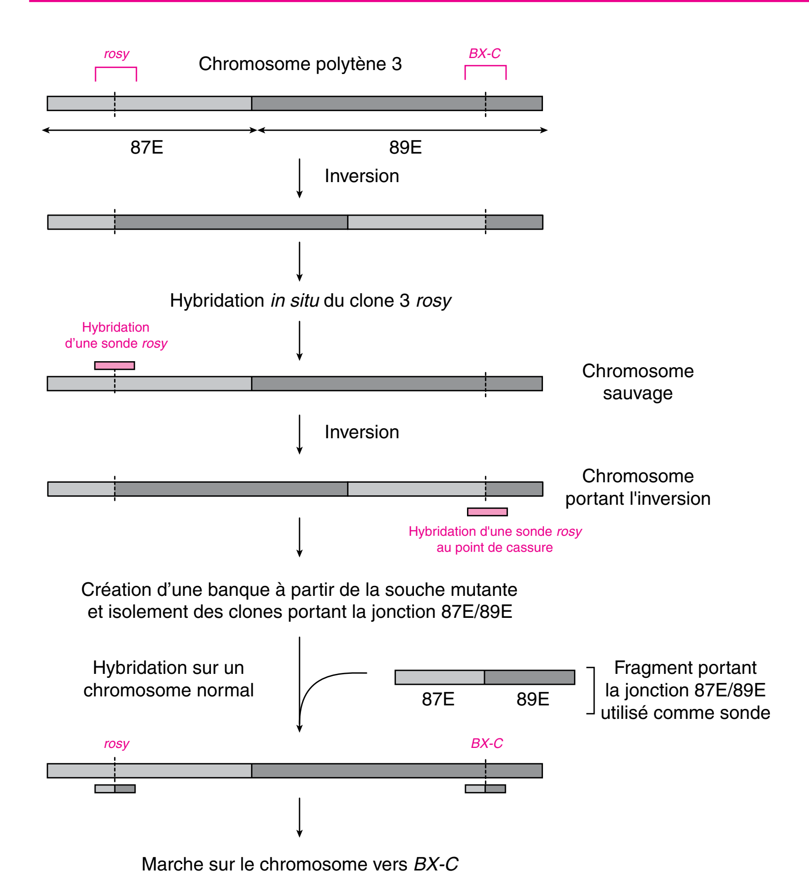


Figure 6.12 Exemple d'utilisation d'une inversion chromosomique identifiée sur un chromosome polytène pour la localisation d'un gène.

Dans l'exemple montré l'inversion d'une partie des régions a permis de rapprocher une région connue portant le gène *rosy* de la région dans laquelle on suspecte la présence d'un gène impliqué dans le développement *BX-C*, mais encore non identifié.

commencer la marche dans cette région. En revanche, ils connaissaient dans la région 87E, la présence du gène *rosy* (impliqué dans la couleur de l'œil) pour lequel ils disposaient d'une sonde spécifique. Des chromosomes issus de mutants présentant une inversion entre les régions 87E et 89E ont été hybridés avec des sondes spécifiques du gène *rosy*. Ainsi, une sonde s'hybrida au niveau du point de cassure. Une banque réalisée à partir de l'ADN présentant l'inversion a été criblée avec la sonde du point de cassure afin d'isoler le fragment situé au point d'inversion qui doit contenir à la fois les séquences 87E



et 89E. Après hybridation sur un chromosome normal, ce fragment révèle bien les deux régions 87E et 89E. Avec la partie du fragment 87E/89E reconnaissant 89E, la marche vers le gène *BX-C* peut commencer.

6.4 CYTOGÉNÉTIQUE ET ASSIGNATION CHROMOSOMIQUE

Cette approche du génome eucaryote s'emploie pour une localisation des gènes et des séquences anonymes sur les chromosomes. En général, ce sont des séquences déjà connues qui, par diverses méthodologies, sont assignées sur un chromosome.

La technique d'hybridation de cellules somatiques est utilisée pour la cartographie du génome de nombreuses espèces. Elle consiste comme le montre la *figure 6.13* à fusionner des cellules d'espèces différentes. La fusion cellulaire (fibroblastes humains/cellules tumorales de souris par exemple) est suivie par une fusion des noyaux, et les cellules hybrides uninucléées contiennent en mélange des chromosomes humains et des chromosomes murins. Comme les chromosomes humains diffèrent des chromosomes murins par leur nombre et leur taille, ils peuvent être distingués au sein d'une cellule hybride. Cependant au cours des divisions cellulaires, les cellules hôtes de souris, immortelles car tumorales, perdent de façon aléatoire une partie des chromosomes humains. Cette élimination peut être suivie en microscopie grâce aux techniques de cytogénétique, technique qui repose sur la coloration des chromosomes. Les cellules hybrides sont cultivées sous forme de lignées séparées, chacune d'elle ayant un contenu distinct mais bien identifié en chromosomes humains.

Avec une banque chromosomique complète, c'est-à-dire dont l'ensemble des lignées contient tous les chromosomes d'une espèce, ici l'homme, il s'avère possible de rechercher la présence d'un gène ou d'un marqueur de type mini ou microsatellite. Comme la séquence est connue, la recherche s'effectue le plus souvent par PCR avec comme matrice l'ADN extrait des différentes lignées. La combinaison des résultats obtenus permet d'assigner celle-ci à un chromosome, comme le montre la comparaison des lignées hybrides présentée dans le *tableau 6.3*.

En comparant les résultats, on observe que :

- la séquence *A* se trouve dans deux lignées 1 et 3, qui ont en commun le chromosome 21 ;
- la séquence *B* est présente dans les lignées 3, 4 et 5, qui toutes possèdent le chromosome 10 ;
- la séquence *C* est mise en évidence dans les lignées 1 et 3, qui ont en commun le chromosome 21 ;

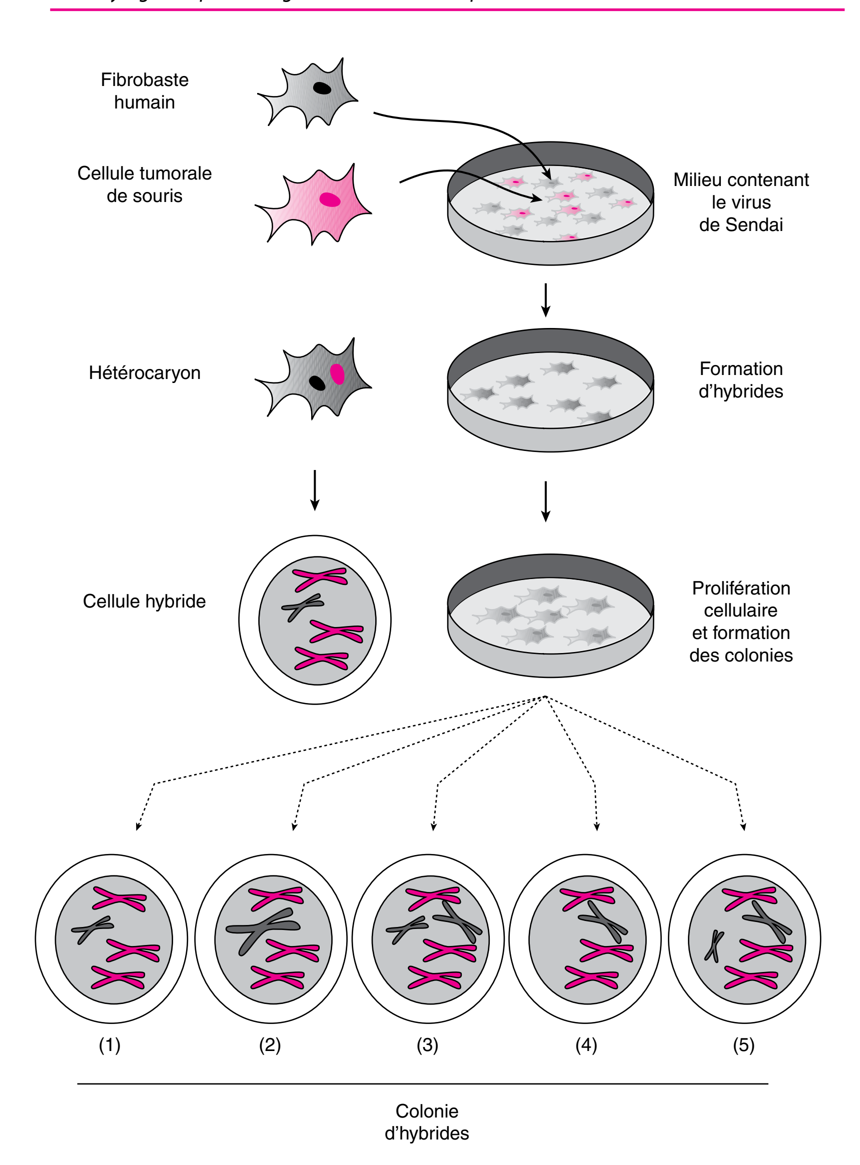


Figure 6.13 Techniques d'obtention d'hybrides somatiques.

Les hybrides (homme/souris) sont produits par fusion cellulaire grâce au virus de Sendai qui est utilisé comme agent de fusion. Dans les hétérocaryons, les deux noyaux vont fusionner, et la cellule hybride contiendra dans son noyau le génome murin et le génome humain. Très rapidement, une partie des chromosomes humains disparaîtront alors que la cellule tumorale conservera la totalité du génome murin. Par analyse cytogénétique, on caractérise chaque colonie hybride par son contenu en chromosomes humains encore présents dans les cellules tumorales. L'identification par cytogénétique a permis d'isoler des colonies hybrides qui diffèrent par leur contenu en chromosomes humains.

TABLEAU 6.3 SÉQUENCES HUMAINES PRÉSENTES DANS LES LIGNÉES HYBRIDES ET CHROMOSOMES HUMAINS CONTENUS DANS CHACUNE D'ELLES.

		Lignées cellulaires hybrides				
		1	2	3	4	5
Séquences humaines recherchées	*A*	+	–	+	–	–
	B	–	–	+	+	+
	C	+	–	+	–	–
	D	+	+	–	+	–
Chromosomes humains présents dans la lignée	1	–	+	–	–	–
	10	–	–	+	+	+
	21	+	–	+	–	–

➤ la séquence *D* est observée dans les lignées 1, 2 et 4, qui en revanche ne possèdent pas le même assortiment chromosomique ; le résultat dans ce cas s'avère inexploitable.

On peut donc assigner au chromosome 10 la séquence *B* et au chromosome 21 les séquences *A* et *C*. De tels résultats amènent à une localisation « grossière » d'une séquence. On détermine le chromosome porteur de la séquence, sans toutefois en préciser la région (bras court, bras long par exemple).

6.5 ANALYSE DE LA LIAISON GÉNÉTIQUE ET TEST DU CHI-DEUX OU χ^2

En génétique, les résultats expérimentaux sont souvent proches des résultats attendus, mais cependant légèrement différents. Un test statistique s'avère nécessaire pour confronter ces résultats entre eux et valider l'hypothèse de départ. Le test du χ^2 permet ce contrôle. Il constitue un moyen de quantifier les différents écarts attendus, liés au seul fait du hasard, entre une valeur prédite et celle obtenue expérimentalement, si l'hypothèse est vraie. Cependant, à toute valeur prise par un écart correspond une probabilité, il s'avère donc nécessaire de définir une valeur seuil au-delà de laquelle l'hypothèse est rejetée. Par convention, une probabilité inférieure ou égale à 5 % est considérée comme un critère de rejet de l'hypothèse. Que signifie cette valeur de probabilité ?

➤ Si l'hypothèse est correcte et du simple fait du hasard, c'est la probabilité d'avoir un écart au moins aussi grand par rapport aux résultats prévus.



- Si la probabilité est supérieure à 5 %, cela signifie que les résultats sont compatibles avec l'hypothèse de départ.
- Le résultat du test du χ^2 dépend forcément du nombre d'individus analysés. Plus il est important, plus le test est fiable.

Le χ^2 est calculé à partir de la formule suivante :

$$\chi^2 = \sum(O - E)^2/E \text{ pour toutes les classes analysées}$$

où : O est égale au nombre observé dans une classe et E (du terme anglais « espected ») celui attendu.

En analyse de liaison génétique, à partir de quel écart à 50 % peut on considérer que deux gènes sont liés ? Le seul critère qui puisse être pris en compte c'est l'absence ou la présence d'un assortiment indépendant. Pour cela, il faut tester l'hypothèse de l'absence de liaison (hypothèse nulle). Si elle est rejetée par un test de χ^2, l'existence d'une liaison génétique sera retenue.

En prenant l'exemple d'une population d'individus obtenus après un croisement-test, on dénombre 600 descendants répartis comme suit :

- 170 de phénotype AB (parental) ;
- 136 de phénotype Ab (recombinant) ;
- 134 de phénotype aB (recombinant) ;
- 160 de phénotype ab (parental).

La fréquence de recombinaison est de 45 %. Ceci pourrait être un cas de liaison génétique car la valeur est inférieure à 50 %. Un test de χ^2 est donc nécessaire pour calculer la probabilité que le résultat obtenu est bien du au hasard. Pour chaque allèle, les proportions sont de :

Ségrégation de *A* et *a*

		A	*a*	Total
Ségrégation de *B* et *b*	*B*	170	134	304
	b	136	160	296
	Total	306	294	600

Les proportions alléliques sont de 306/600 pour *A*, 294/600 pour *a*, 304/600 pour *B* et 296/600 pour *b*. À partir de ces proportions, les valeurs attendues (E) pour les individus *AB* sont de : (306/600 × 304/600) × 600 = 155,04.



Le tableau des valeurs attendues est :

		Ségrégation de *A* et *a*		
		A	*a*	Total
Ségrégation de *B* et *b*	B	155,04	148,96	304
	b	150,96	145,04	296
	Total	306	294	600

La valeur de χ^2 est de :

Génotype	O	E	$(O - E)^2/E$
AB	170	155,04	1,44
Ab	136	150,96	1,48
aB	134	148,96	1,50
ab	160	145,04	1,54
Total	306	294	$\chi^2 = 5,97$

La valeur du χ^2 est utilisée pour trouver la probabilité (*p*) associée au test, à l'aide d'une table standardisée (non présentée) de χ^2. De manière générale, dans un test statistique, le degré de liberté (dl) correspond au nombre de valeurs indépendantes. Dans le tableau de données ci-dessus les totaux des lignes et des colonnes proviennent de résultats expérimentaux. Aussi, connaître n'importe quelle valeur du tableau permet d'en déduire les autres. Le dl est donc de 1. De manière globale, le dl est donné par le nombre de classes dans une ligne moins un multiplié par le nombre de classes dans une colonne moins un. Dans l'exemple développé ici, la valeur de 5,97 est largement inférieure au seuil de probabilité de 5 % donné dans la table du χ^2. Ainsi, l'hypothèse d'un assortiment indépendant (hypothèse nulle) doit être rejetée. On en conclut que les loci sont liés.

6.6 HÉRÉDITÉ LIÉE AU SEXE

De nombreux végétaux ainsi que la plupart des animaux présentent un dimorphisme sexuel, et dans la majorité des cas le sexe est déterminé par une paire de **chromosomes sexuels**. Chez les humains, les cellules contiennent 22 paires de chromosomes homologues, les **autosomes**, et une paire de chromosomes sexuels, identiques chez la femme (2 chromosomes X) et non identiques chez l'homme (1 chromosome X



et 1 chromosome Y). Lors de la méiose, chez la femelle les chromosomes sexuels s'apparient comme les autosomes *(Fig. 6.14)* et se disjoignent de sorte que chaque ovule, reçoit un chromosome X. On dit que la femme est de sexe **homogamétique**. En revanche chez l'homme, les chromosomes X et Y s'apparient sur de courtes régions homologues, les **régions pseudo-autosomales**, et lors de la ségrégation donnent deux types de spermatozoïdes, une partie recevant le X et l'autre le Y. De ce fait, l'homme est de sexe **hétérogamétique**. Les régions non homologues de ces chromosomes sont appelées **régions différentielles**.

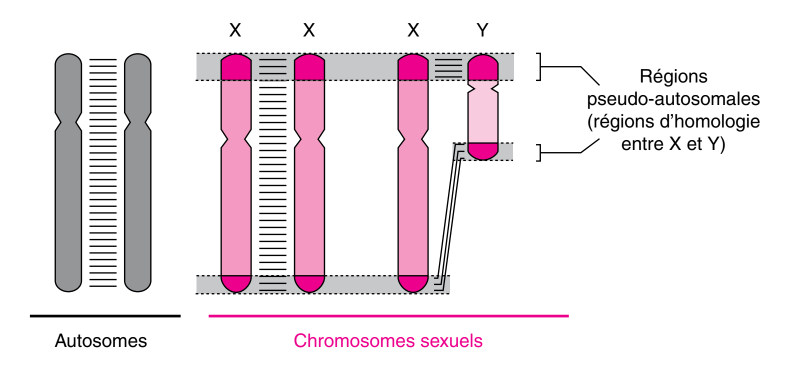


Figure 6.14 Région d'appariement des chromosomes sexuels chez l'Homme.

Alors que les autosomes et les chromosomes X peuvent s'apparier sur toute leur longueur, seules les courtes régions télomériques du X et du Y peuvent s'apparier au cours de méiose. Les régions pseudo-autosomales ont été identifiées en observant l'appariement des chromosomes X et Y lors de la méiose.

Déterminisme du sexe

La détermination du sexe est génétiquement définie. En prenant l'homme comme exemple, un gène, ***SRY*** pour « Sex Regulation on the Y », localisé sur le bras court du chromosome Y, joue un rôle crucial dans le déterminisme du sexe. En effet, il a été montré que l'inactivation de ce gène chez un individu XY aboutit à un phénotype féminin. *SRY* assure une fonction d'interrupteur et le produit de son expression est un facteur de transcription qui se lie sur l'ADN dans les gonades indifférenciées les transformant en testicules ; en revanche, en son absence ces mêmes gonades deviendront des ovaires. Après formation des testicules, il y a activation des gènes impliqués dans la synthèse et sécrétion de testostérone, hormone liposoluble responsable des caractères sexuels secondaires masculins (forme corporelle, pilosité, etc.) propres au phénotype masculin. La testostérone se fixe sur son récepteur, codé par le gène *AR* « Androgen Receptor » et porté par le chromosome X.



Le complexe testostérone-AR passe dans le noyau où il agit sur des facteurs de transcriptions qui activent la transcription de gènes de la « masculinité ». En l'absence du gène *AR*, la testostérone ne peut pas agir. L'individu présente un syndrome d'insensibilité aux androgènes, dont fait partie la testostérone, et il n'aura aucun caractère masculin.

Hérédité liée au chromosome X

Le chromosome X contient plusieurs centaines de gènes, très peu étant associés à la fonction sexuelle. Un grand nombre de ces gènes se trouvent dans la région différentielle et n'ont donc pas d'équivalent sur le chromosome Y. Ces gènes présentent un mode de transmission héréditaire dite transmission liée à l'X. Ce mode de transmission a été mis en évidence en effectuant des croisements réciproques entre des drosophiles à yeux rouges et à yeux blancs :

- En croisant des mâles à yeux blancs avec des femelles à yeux rouges, 100 % des descendants de la F1 ont les yeux rouges. Les individus de la F1 croisés entre eux donnent une population F2 où 100 % des femelles ont les yeux rouges alors que chez les mâles, 50 % ont les yeux rouges et 50 % les yeux blancs.
- En effectuant le croisement réciproque, c'est-à-dire des mâles à yeux rouges avec des femelles à yeux blancs, au niveau de la F1 toutes les femelles ont les yeux rouges et tous les mâles les yeux blancs. En croisant les individus de la F1 entre eux, on obtient pour les mâles et les femelles 50 % de drosophiles à yeux rouges et 50 % à yeux blancs.

Ce mode de transmission s'explique par la présence d'un gène de la région différentielle du chromosome X. L'allèle sauvage *R* est dominant et son produit est responsable de la couleur rouge, et l'allèle muté *r* est récessif, couleur blanche de l'œil. Dans le premier croisement les femelles sont *RR* et les mâles *r* (un seul chromosome X). Tous les descendants de la F1 héritent d'un chromosome X maternel porteur de l'allèle *R*, ils ont donc tous les yeux rouges. En revanche, lors du croisement réciproque, les femelles sont *rr* et les mâles *R*. Les femelles de la F1 ayant héritées du chromosome X paternel auront toutes les yeux rouges alors que les mâles héritent d'un X maternel, porteur de l'allèle récessif, et auront les yeux blancs. Ceci permet de comprendre les résultats obtenus à la F2. Pour le premier croisement :

F1 : ♀ *Rr* × ♂ *R*

F2 : ♀ *Rr* ou *RR* donc 100 % yeux rouges

♂ *r* ou *R* d'où 50 % de chaque phénotype



Les rapports phénotypiques différents pour les descendants des deux sexes constituent une caractéristique de l'hérédité liée au sexe. Du fait des régions différentielles, des mutations affectant des gènes présents dans cette région auront, même si elles conduisent à des allèles récessifs, des répercutions sur la descendance. Ainsi, dans les familles atteintes de pathologies liées à l'X on observe un déséquilibre au niveau des arbres généalogiques. Plus d'individus mâles sont atteints, l'exemple le plus connu étant l'arbre généalogique de la reine Victoria et la transmission de l'hémophilie A à ses descendants. D'autres pathologies sont dues à des mutations récessives du X comme la myopathie de Duchenne (mutation du gène codant la dystrophine), le syndrome de féminisation testiculaire (mutation d'un récepteur aux androgènes), le daltonisme (mutation au niveau d'un photo récepteur), etc.

6.7 TRANSMISSION DE TRANSGÈNES

L'étude des génomes des organismes complexes se révèle d'une extrême importance. L'objectif final étant l'étude approfondie du génome humain en vue d'appréhender les fonctions telles que les capacités cognitives, le comportement social, etc. En ce qui concerne l'étude de la fonctionnalité des gènes humains, on comprend facilement que l'éthique n'autorise pas de mener ces études directement sur l'être humain. Les expériences effectuées aujourd'hui concernent exclusivement des lignées cellulaires humaines cultivées *in vitro* et dans une moindre mesure la participation d'individus volontaires. Le choix d'un modèle dont la physiologie est proche de celle de l'homme s'est donc révélé important afin d'espérer entreprendre ce type d'études.

À côté des vertébrés inférieurs (poissons et batraciens), le mammifère qui a retenu l'attention des généticiens est la souris. En effet, ses caractéristiques sont intéressantes : sa petite taille facilite son élevage, une femelle donne naissance à une portée de 8 à 12 petits, la durée de gestation est de 3 semaines et les jeunes atteignent la maturité sexuelle à l'âge de 6 semaines. La souris possède en particulier une structure cérébrale complexe et un système immunitaire proches de ceux de l'homme. Ce petit rongeur est utilisé depuis des décennies par les généticiens. Ainsi des dizaines de mutants ont, par le passé, pu être caractérisés et les gènes impliqués ont été localisés sur le génome. La plupart d'entre eux ont été clonés et les fonctions, qui y sont associées ont été déterminées. Dans une démarche de **génétique inverse**, une technique d'invalidation de gène, mettant à profit la recombinaison homologue, a pu être mise au point chez la souris *(Fig. 6.15)*. Les souris qui possèdent un gène invalidé sont aussi appelées souris « **knock-out** » ou **mutants nuls**.



Figure 6.15 Méthodologie générale pour la création de souris transgéniques « knock-out ».

On constate qu'il faut trois générations de souris, à partir de la mère, pour obtenir les individus homozygotes pour le gène invalidé (invalidé/invalidé). Les cellules ES utilisées pour la transformation sont issues d'une lignée de souris agouti (couleur du pelage marron). Le suivi de ce caractère permet aux généticiens de sélectionner plus facilement les souris chimères portant le gène invalidé en F1 et les souris hétérozygotes en F2.



La technique est basée sur l'utilisation de cellules souches embryonnaires (voir dans la même collection *Biologie cellulaire*). Le gène est inactivé par une stratégie de recombinaison homologue proche de celle utilisée chez la levure (voir *figure 4.14*). Dans ce cas, le gène d'intérêt est remplacé par un marqueur de sélection constitué d'un gène bactérien qui confère la résistance à un antibiotique, la néomycine, auquel les cellules ES sont sensibles. Le gène ainsi invalidé est introduit grâce à un plasmide dans les cellules ES par **électroporation** et les cellules transfectées sont sélectionnées en présence de néomycine. Les cellules ES résistantes à la néomycine sont hétérozygotes, car elles possèdent un allèle normal et un allèle invalidé. L'allèle normal étant dominant, les cellules se développent normalement. Les cellules ES, ainsi modifiées, sont ensuite introduites, par microinjection, dans des embryons de souris au stade blastocyste. Les embryons sont donc constitués de cellules ES normales et de cellules ES portant le gène invalidé. Ils sont ensuite implantés dans l'utérus d'une femelle pseudo-gestante. Celle ci va donner naissance à des **souris chimères**, qui posséderont certains tissus dérivant des cellules ES normales et d'autres provenant des cellules ES portant le gène invalidé. Parmi les souris chimères, les généticiens doivent sélectionner celles dont les cellules de la **lignée germinale** dérivent des cellules ES portant le gène invalidé. Il est en effet indispensable de transmettre le gène invalidé aux générations suivantes. Ainsi, une souris chimère est croisée avec une souris normale, afin de rechercher dans la descendance les souris hétérozygotes pour le gène invalidé. Finalement, pour obtenir les souris « knock-out », il sera nécessaire de croiser deux souris hétérozygotes pour le gène invalidé qui donneront une descendance composée à 50 % de souris hétérozygotes, 25 % de souris homozygotes normales et 25 % de souris « knock-out », homozygotes pour le gène invalidé. Comme nous l'avons vu précédemment chez la levure, certains gènes sont essentiels à la viabilité. Chez la levure, le nématode ou bien la souris, un gène sur trois est essentiel à la viabilité. Il n'est donc pas possible d'obtenir une souris « knock-out » qui porte une invalidation pour un gène essentiel. Afin de contourner en partie ce problème et permettre tout de même l'étude de ces gènes, une stratégie originale à été appliquée aux souris « knock-out ». Il s'agit de l'**invalidation conditionnelle** des gènes. Cette technique met à profit le système Cre/Lox *(Fig. 6.16)*.

L'enzyme Cre est une recombinase du bactériophage P1. Elle catalyse la recombinaison entre les deux séquences d'ADN de 34 pb appelées *LoxP*. Si un fragment d'ADN est placé entre deux séquences LoxP, la protéine Cre provoquera l'excision de ce fragment d'ADN et d'une séquence *LoxP*. L'utilisation de ce système *in vivo* nécessite le

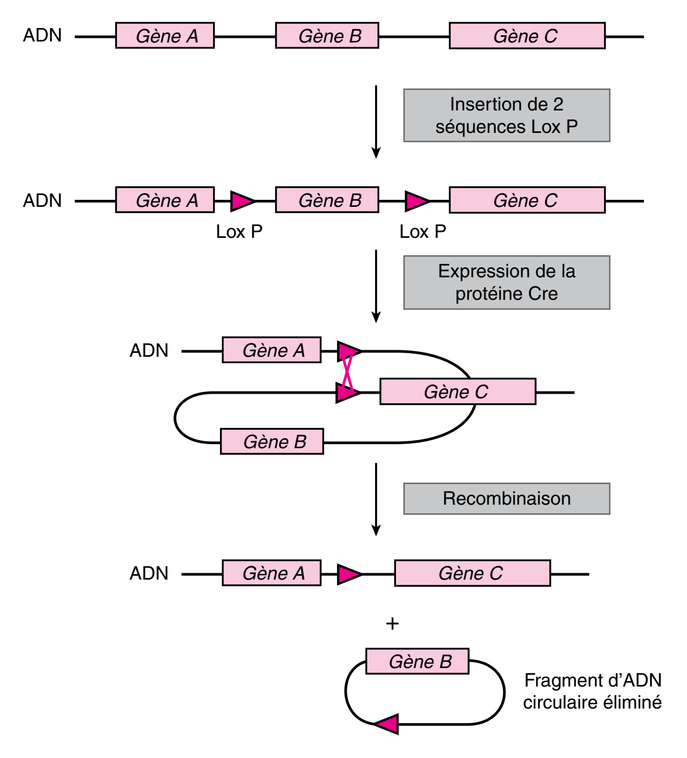


Figure 6.16 Invalidation conditionnelle par le système Cre/Lox.

Deux séquences Lox P sont insérées dans le chromosome de la souris de part et d'autre du fragment à déléter (le gène B dans cet exemple). L'expression ciblée de la recombinase Cre permet l'excision du fragment par recombinaison homologue entre les deux séquences Lox P. Le fragment d'ADN circulaire est éliminé car il ne possède pas de centromère.

croisement de deux lignées de souris transgéniques. Une première lignée exprime le gène *Cre* et la seconde le gène à invalider, flanqué d'une séquence *LoxP* à chacune de ses extrémités. Il est possible de récupérer, dans la descendance, des souris qui portent les deux types de transgènes. L'invalidation du gène d'intérêt va donc dépendre de l'expression de la protéine Cre. Il est donc nécessaire de contrôler cette expression au cours du développement et/ou en fonction des tissus. Ceci est réalisé en utilisant, pour l'expression de la recombinase, des promoteurs adaptés qui pourront être activés à certains stades du développement ou dans certains tissus. Ainsi, l'expression de la protéine Cre conduira à l'invalidation du gène dans le tissu ou l'organe ciblé au moment choisi par le généticien.



6.8 DÉRIVES AUX LOIS DE MENDEL DANS LA TRANSMISSION DES CARACTÈRES

L'épistasie

Le phénotype d'un individu résulte pour de nombreux caractères de l'interaction entre gènes. Les effets d'un gène ne peuvent être vus en présence d'un allèle particulier d'un autre gène. On parle d'**épistasie** pour ce phénomène où tout type d'interactions entre gènes entraîne une modification du rapport 9:3:3:1 de la génération F2. De manière générale, on dit que l'expression d'un gène masque l'expression de l'autre. Les phénomènes épistatiques se produisent uniquement lorsque les gènes sont impliqués dans la même voie biochimique.

À titre d'exemple chez la souris, des gènes impliqués dans la couleur du pelage peuvent interagir entre eux et modifient la couleur du pelage. Des souris noires de race pure ont pour génotype *BBCC* où *B* code pour un pigment noir et le produit de *C* permet le dépôt de pigment. Les allèles récessifs *b* et *c* correspondent respectivement à des pigments bruns et à l'absence de dépôts de pigments. Le croisement de souris noires avec des souris blanches de génotype *bbcc*, donne des descendants F1 noirs et hétérozygotes : *BbCc*. En F2 pour un même phénotype, divers génotypes sont possibles :

- animaux noirs : *BBCC*, *BBCc*, *BbCC*, *BbCc* dans un rapport de 9/16 ;
- animaux bruns : *bbCC*, *bbCc* dans un rapport de 3/16 ;
- animaux blancs : *BBcc*, *Bbcc*, *Bbcc* dans un rapport de 4/16 ; la couleur étant due ici à l'absence de dépôt de pigment liée à la présence des deux allèles *cc*.

On parle ici d'épistasie récessive. En supposant maintenant que l'allèle *C* dominant empêche le dépôt de pigment, les rapports seraient là encore modifiés :

- animaux noirs : *BBcc*, *Bbcc* dans un rapport de 3/16 ;
- animaux bruns : *bbcc* dans un rapport de 1/16 ;
- animaux blancs : *BBCC*, *BBCc*, *BbCC*, *BbCc*, *bbCC* dans un rapport de 12/16.

La présence de l'allèle *C*, dominant, entraîne l'absence de pigmentation et impose la coloration blanche du pelage ; on parle d'épistasie dominante.

Les interactions génétiques ont donc des effets sur la ségrégation des caractères. De plus, les gènes impliqués dans une cascade métabolique ou biochimique peuvent être ordonnés si pour ces gènes, on dispose de mutations donnant des effets différents. Si le gène épistatique est en amont dans la voie de synthèse, on constate que, par exemple, dans le déterminisme des groupes sanguins, la mutation du gène *H* (inactivation d'une fucosyltransférase, allèle récessif *h*) interdit la



synthèse des antigènes A et B. Pourtant chez les individus porteurs de cette mutation (individus dits « Bombay » découverts pour la première fois en Inde) le gène *I* responsable de la synthèse des motifs biochimiques caractéristiques des groupes A, B et AB, est de type sauvage et fonctionnel. C'est donc bien une mutation inactivatrice du gène *H* qui impose le phénotype « Bombay » et non des mutations du gène *I* (voir chapitre 1).

Si le gène épistatique est en aval dans la voie de synthèse, comme dans le cas de *C* dominant, on constate que malgré la synthèse du pigment son dépôt est inhibé, d'où l'absence de pigmentation des souris portant cet allèle.

Les gènes suppresseurs

La suppression correspond à un autre type d'interaction entre gènes. Un **suppresseur** est l'allèle mutant d'un gène qui inverse l'effet de la mutation d'un autre gène en redonnant un phénotype sauvage. Différents événements peuvent conduire à la suppression. Une mutation peut intervenir dans le gène muté et restaure sa fonctionnalité, on parle de suppresseur intragénique ; une mutation peut affecter un autre gène que le gène muté et restaurer également le phénotype sauvage, il s'agit d'un suppresseur extragénique. Les suppresseurs extragéniques ne modifieront le phénotype que chez les doubles mutants.

À titre d'exemple chez la drosophile une mutation au niveau du gène *Pd* conduit à un allèle récessif *pd* qui à l'état homozygote modifie la couleur de l'œil en l'absence de suppression. Un allèle muté, *su*, du gène *Su*, ne génère pas en lui-même le phénotype détectable, mais en revanche modifie l'effet de l'allèle *pd*. Si on croise des individus *PdPdsusu* (yeux rouges) avec des individus *pdpd SuSu* (yeux violets), la F1 est double hétérozygote. Les individus de la F1 croisés entre eux vont donner des descendants de phénotypes :

- Yeux rouges dans un rapport de 13/16 ; ce phénotype correspond aux individus *PdPdSuSu*, *PdpdSuSu*, *PdPd Susu*, *PdpdSusu*, *PdPdsusu*, *Pdpdsusu*, et *pdpdsusu*. Chez ces derniers la présence du suppresseur *su* corrige l'effet de la mutation pd.
- Yeux violets dans un rapport 3/16 apporté par les individus de génotype : *pdpdSusu*, *pdpd SuSu*.

Dans l'exemple choisi, on a un suppresseur récessif qui agit sur une mutation récessive. Il existe des suppresseurs dominants et récessifs qui peuvent agir sur des mutations dominantes ou récessives. Bien que l'on puisse confondre la suppression avec l'épistasie, les rapports obtenus en F2 diffèrent et permettent de distinguer les deux phénomènes.





POINTS CLEFS

- La mitose concerne à la fois les cellules haploïdes et diploïdes. Elle aboutit à deux cellules filles génétiquement identiques à la cellule mère.
- La méiose ne peut exister que pour des cellules obligatoirement diploïdes et aboutit à partir d'une cellule mère diploïde à quatre cellules haploïdes génétiquement différentes entre elles.
- La recombinaison méiotique désigne tout processus méiotique qui donne un produit haploïde présentant une combinaison d'allèles différente de celle portée par les génotypes haploïdes à l'origine du méiocyte**.**
- L'analyse de la descendance à l'issue d'un croisement test ou « test-cross » permet d'identifier les recombinants dont les phénotypes diffèrent de ceux des parents P1.
- Les recombinants sont produits par deux processus distincts : l'assortiment indépendant et les crossing-over.
- Le brassage interchromosomique correspond à une redistribution aléatoire des chromosomes parentaux lors de l'anaphase de la première division méiotique.
- La recombinaison intrachromosomique accroît encore un peu plus la diversité génétique des populations ultérieures.
- La fréquence de recombinaison entre deux loci reflète la distance génétique qui les sépare sur le même chromosome.
- L'interférence traduit le fait qu'un crossing-over diminue la probabilité qu'un deuxième crossing-over se produise dans une région très proche.
- Les ascospores sont réparties dans un asque linéaire en fonction du profil de ségrégation résultant des deux divisions méiotiques. Si un crossing-over s'effectue entre un locus et un centromère, les ascospores seront réparties par bloc de deux en fonction des chromatides impliquées dans la recombinaison. On peut ainsi réaliser la cartographie des centromères.
- Des mutations facilement identifiables comme des inversions sur un chromosome polytène contribuent à la détermination de la position approximative de l'allèle muté responsable d'un nouveau phénotype.
- Le test du χ_2 constitue un moyen de quantifier les différents écarts attendus, liés au seul fait du hasard, entre une valeur prédite et celle obtenue expérimentalement, si l'hypothèse est vraie.
- Les chromosomes X et Y s'apparient sur de courtes régions homologues, les régions pseudo-autosomales ; les régions non homologues de ces chromosomes sont appelées régions différentielles.
- Le gène *SRY*, « Sex Regulation on the Y », localisé sur le bras court du chromosome Y, joue un rôle crucial dans le déterminisme du sexe masculin.



- Un grand nombre des gènes portés par le chromosome X se trouve dans la région différentielle, ils n'ont donc pas d'équivalent sur le chromosome Y. Ils présentent un mode de transmission héréditaire dite transmission liée à l'X.
- Les rapports phénotypiques différents pour les descendants des deux sexes, observés après des croisements réciproques, constituent une caractéristique de l'hérédité liée au sexe.
- Dans une démarche de génétique inverse, une technique d'invalidation de gène, mettant à profit la recombinaison homologue, a pu être mise au point chez la souris. Les souris qui possèdent deux copies d'un gène invalidé sont appelées souris « knock-out » ou mutants nuls.
- L'épistasie traduit la situation dans laquelle l'expression phénotypique différentielle d'un génotype à un locus est sous la dépendance d'un autre locus.
- Un suppresseur est l'allèle mutant d'un gène qui inverse l'effet de la mutation d'un autre gène en redonnant un phénotype sauvage.

QCM - QROC

6.1 Quelle affirmation est exacte ?

a) Par division mitotique, une cellule génère deux cellules filles génétiquement différentes.

b) La mitose est réservée exclusivement aux cellules diploïdes.

c) Au cours de la méiose, il y deux réplications de l'ADN puis une seule division cellulaire.

d) La méiose est à l'origine de la diversité génétique des individus d'une même espèce.

6.2 Les affirmations suivantes sont-elles vraies ou fausses ?

a) Au cours de la prophase de la 1re division méiotique les chromosomes sexuels chez l'homme s'apparient sur toute leur longueur.

b) Au cours de la prophase de la 1re division méiotique les chromosomes sexuels chez la femme s'apparient sur toute leur longueur.

c) La division réductionnelle concerne la séparation des chromatides d'un chromosome, et la division équationnelle la séparation des chromosomes homologues.

d) Les crossing-over s'effectuent entre chromosomes homologues lors de la 1re division méiotique.

6.3 Parmi les propositions suivantes, quelles sont celles qui sont vraies ?

a) Les mécanismes de recombinaison génétique peuvent générer plus de 50 % de recombinants.



b) Deux gènes liés sont portés par des chromosomes différents.

c) Les fréquences de recombinaison entre deux loci permettent d'estimer la distance qui les sépare sur un chromosome.

d) La distribution des ascospores de même génotype dans un asque linéaire est toujours identique.

e) La présence de doubles crossing-over dans une même région interfère sur le calcul des distances génétiques entre les caractères étudiés.

f) Le test du χ^2 valide, sur la base d'une hypothèse vraie, la part du hasard dans les distributions phénotypiques observées par rapport à celle attendues.

6.4 Chez les humains, lorsque des caractères sont portés par la région différentielle du chromosome X :

a) Le père transmettra les allèles de ces caractères à ses fils.

b) Si une femme, porteuse saine, se marie avec un homme sain, la probabilité d'avoir un fils normal est de 25 %.

c) Pour les caractères analysés, le phénotype des filles dépendra exclusivement du génotype maternel.

d) Si une mutation affecte l'un des caractères, le phénotype qui en résulte sera majoritairement observé chez les descendants mâles.

Quelles affirmations pouvez-vous retenir ?

6.5 Retenir les propositions correctes.

a) Les interactions géniques peuvent masquer des phénotypes.

b) Les gènes épistatiques masquent les phénotypes.

c) Dans le cas d'épistasie dominante, par croisement entre individus d'une F1, une forte partie d'entre eux sera de phénotype sauvage.

d) Les gènes suppresseurs ont toujours pour cibles les allèles récessifs d'autres gènes.

6.6 Si sur des individus *AaBb* d'une F1, issue d'un croisement *AABB* × *aabb*, on effectue un croisement test, au niveau de la descendance quel pourcentage sera *Aabb* si :

a) les gènes sont non liés ;

b) les gènes sont liés et distants de 20 cM ;

c) trop proche pour qu'il y ait un crossing-over.

6.7 Un allèle récessif *h*, porté par le X, est responsable de l'hémophilie, une femme saine dont la mère est hémophile épouse un homme normal. Quelle est la probabilité d'avoir un fils hémophile ?

6.8 La synthèse de pigments rouges chez les fleurs se fait en deux étapes :

gène P *gène Q*

Pigments blancs → Pigments magenta → Pigments rouges

Mutés les gènes P et Q codent des produits incapables d'assurer pour l'allèle *p* la synthèse de pigment magenta et *q* celle de pigment rouge. Pour les deux gènes les mutations sont récessives.

a) Quelle sera la couleur de la fleur d'une plante de génotype *PPqq*, et celle d'une plante de génotype *ppQQ*.

b) À quels rapports phénotypiques doit-on s'attendre en croisant des plantes *QqPp* entre elles ?

RÉPONSES

6.1 d) Vrai ; **b)** les organismes haploïdes comme la levure se divisent par mitose

6.2 b) et **d)** vraies

6.3 c), **e)** et **f)** vraies ; **a)** il ne peut jamais y avoir plus de 50 % de recombinants ; **d)** la distribution des ascospores peut varier si des crossing-over s'effectuent entre le centromère et un locus donné.

6.4 b) et **d)** vraies ; **a)** faux car le père ne peut pas transmettre de chromosome X à ses fils ; **c)** faux, si la mère est hétérozygote *Aa* et le père *a*, les filles auront deux génotypes possibles *Aa* ou *aa* qui donneront des phénotypes distincts.

6.5 a) et **b)** vraies ; **c)** faux : la majeure partie sera de phénotype muté imposé par la présence de l'allèle dominant du gène à effet épistatique ; **d)** faux : ils peuvent également agir sur des allèles sauvages.

6.6 a) 25 % des gamètes de la F1 seront de génotype *Ab*, la contribution de l'autre parent lors d'un croisement test étant *ab*, 25 % des individus auront le génotype *Aabb* ; **b)** la distance qui sépare les deux caractères étant déduite des fréquences de recombinaison, au niveau de la F1 les recombinants possibles sont *aB* et *Ab*, et représentent au total 20 % de la descendance. Donc, la moitié des recombinants, soit 10 % des individus de la population analysée seront *Aabb* ; **c)** aucun individu ayant ce phénotype ne sera observé du fait de l'absence de crossing-over.

6.7 25 % : la mère est hétérozygote (saine mais de mère hémophile), donc *Hh* (*H* désignant l'allèle sauvage dominant). Il y a une chance sur



deux qu'elle transmette l'allèle *h* et également une chance sur deux que l'enfant soit un fils ; 1/2 × 1/2 = 1/4 soit 25 %.

6.8 a) Une plante de génotype *PPqq* ne pourra pas transformer les pigments magenta en pigments rouges du fait de la mutation de l'allèle *q* présent à l'état homozygote. Les fleurs seront de couleur magenta. Celle de génotype *ppQQ*, malgré la fonctionnalité du gène *Q*, n'aura aucun pigment magenta, intermédiaire dans la voie de synthèse, donc les fleurs seront blanches.

b) 9 : 4 : 3. En faisant le tableau de croisement on en déduit : 9/16 de plantes à fleurs rouges (1/16 *PPQQ* + 2/16 *PPQq* + 5/16 *PpQq* + 1/16 *PpQQ*) ; 4/16 de plantes à fleurs blanches (1/16 ppqq + 1/16 ppQQ + 2/16 ppQq) ; 3/16 de plantes à fleurs magenta (1/16 *PPqq* + 2/16 *Pqqq*). Nous sommes ici dans un cas d'épistasie récessive, où l'absence d'un produit fonctionnel en amont de la voie de synthèse bloque la formation des pigments.

CHAPITRE 7

Génétique de l'évolution et du développement des organismes

PLAN

OBJECTIFS

- Définir la notion d'évolution
- Établir les relations entre l'évolution et la génétique du développement des organismes
- Aborder l'origine et l'évolution de la structure des gènes
- Aborder les mutations dans la régulation des gènes comme source de nouveauté évolutive

7.1 GÉNÉTIQUE DE L'ÉVOLUTION

L'évolution selon Darwin repose sur la sélection naturelle. Depuis sa formulation en 1859, cette expression toute simple a fait couler beaucoup d'encre. Sa compréhension contemporaine, défendue par la plupart des biologistes, s'appuie sur le constat que les seuls individus qui survivent (qui sont sélectionnés) sont ceux qui, dans un environnement changeant, sont les plus aptes à se reproduire parce qu'ils y sont les mieux adaptés. L'environnement exerce donc un tri, une sélection qui provoque une évolution en faisant apparaître des adaptations. La sélection peut s'exercer à différents niveaux d'organisation des êtres vivants : celui du matériel génétique, acides nucléiques et protéines ainsi que leurs complexes géniques et chromatiniens et ceux de la cellule, de l'organisme ou de la population. La sélection procède selon un tri, un échantillonnage, caractérisé par son caractère aléatoire. En effet, parmi les innombrables potentialités théoriques offertes par la combinatoire des variations (mutations) de tous types, seul un petit nombre est conservé à chaque génération et dans ce processus ou dérive génétique, le hasard joue un rôle essentiel.



Macro et microévolution : aperçu général et définitions

D'un strict point de vue génétique, les deux mécanismes importants qui alimentent les changements de fréquence des allèles au sein d'une population sont : (i) la dérive génétique (le tri aléatoire) et (ii) la sélection reproductive liée à l'environnement. Même en l'absence de sélection, le tri aléatoire des mutations transmises entre les générations conduira progressivement à ce que la fréquence d'un allèle tende vers 100 % (allèle fixé) ou 0 % (disparition de l'allèle) et ce d'autant plus efficacement que la population sera de petite taille. Pour des populations de plus grande taille, c'est le mécanisme de sélection naturelle qui prédominera. Quelle qu'en soit la cause, le changement de taille d'une population aura donc un effet important sur sa capacité à évoluer par sélection naturelle. Celle-ci suppose en effet :

- qu'une variabilité héritable naturelle existe au sein de la population considérée ;
- que les individus engendrent plus de descendants qu'il n'en peut survivre ;
- qu'au sein de la population, les individus varient dans leur capacité à survivre et à se reproduire ;
- que les reproducteurs les plus actifs transmettent leurs caractéristiques génétiques à leurs descendants alors que les moins actifs ne le font pas.

Si ces diverses caractéristiques contribuent à augmenter l'adaptabilité évolutive des individus qui les possèdent, ils survivront et se reproduiront plus efficacement que les autres. Inversement les caractéristiques qui amoindriront l'adaptabilité des individus qui en sont porteurs contribueront à raréfier la représentation de ces derniers au sein de la population.

On désigne par le terme de **macroévolution** celle qui intervient au niveau de l'espèce ou à un niveau supérieur. Typiquement la macroévolution fait référence aux changements qui modifient l'embryogenèse, les plans d'organisation, les principales caractéristiques des individus et qui sont à la base de la phylogénie des grands groupes du monde vivant. La **microévolution** par contre désigne au sein d'une population des changements évolutifs de moindre ampleur comme ceux qui sont relatifs à la fréquence des allèles. Les biologistes considèrent que macroévolution et microévolution sont liées par le fait que l'accumulation graduelle de petits changements entraîne les changements de plus grande ampleur, même si des sauts évolutifs rapides sont également visibles dans les traces paléontologiques.

Certains biologistes, peu nombreux mais très actifs et relayés par des mouvements religieux, surtout aux USA, admettent l'existence des petits changements évolutifs, mais rejettent l'idée d'une macroévolution qui

s'oppose, à leurs yeux, à l'existence d'un **dessein intelligent** (« intelligent design ») qui n'est, à vrai dire, que la forme moderne de la création *ex nihilo* du monde vivant suivant les dogmes des principales religions monothéistes. Ce mouvement antiévolutionniste fait écho en ce début de XXI[e] siècle à la montée des fondamentalismes religieux radicaux.

Les racines d'une nouvelle discipline : l'évo-dévo

L'étude des processus génétiques intervenant lors du développement des organismes a donné naissance à un courant scientifique très fécond qui s'efforce d'établir un lien entre l'origine des changements observés au cours du développement des organismes depuis l'œuf fécondé jusqu'à l'adulte et l'évolution des gènes conduisant à l'apparition des espèces. Cette discipline bénéficie aujourd'hui des progrès considérables découlant (i) de la découverte des gènes contrôlant notamment le développement embryonnaire précoce et (ii) des connaissances de plus en plus nombreuses apportées par le séquençage complet des génomes.

L'évolution du développement biologique, ou en abrégé l'**évo-dévo**, se propose donc de comparer les processus développementaux des organismes vivants (animaux et des plantes). Elle essaie de déterminer les relations ancestrales qui existent entre les organismes et comment ces processus ont évolué. L'évo-dévo aborde ainsi l'origine et l'évolution du développement embryonnaire. Elle analyse comment les changements dans le développement et les processus qui le dirigent ont produit de nouvelles caractéristiques biologiques. Elle s'intéresse au rôle de la plasticité développementale dans l'évolution et comment l'écologie intervient dans ces phénomènes.

Bien que la vieille théorie d'Ernest Haeckel qui affirmait au XIX[e] siècle, que « l'ontogénie récapitule la phylogénie » – c'est-à-dire que le développement d'un organisme reflète exactement l'évolution de son espèce – soit depuis longtemps rejetée, certains biologistes admettent cependant l'existence de nombreuses filiations évolutives entre ontogénie et phylogénie. Les mêmes processus généraux agiraient dans l'un et l'autre cas comme, par exemple, la réutilisation d'anciens gènes et réseaux de gènes pour créer de nouvelles structures et de nouveaux réseaux. Une sorte de recyclage des instructions géniques qui ressemble à un bricolage (« faire du neuf avec de l'ancien ») selon l'expression de F. Jacob. Le principal intérêt de cette démarche consistait à l'origine à mettre en évidence l'homologie des mécanismes moléculaires et cellulaires qui contrôlent la construction des organismes puis le développement. Aujourd'hui la démarche s'applique à tous les aspects et échelles de l'évolution.



7.2 UNE BRÈVE HISTOIRE DE L'ORIGINE DES GÈNES

L'ARN a-t-il précédé l'ADN comme support moléculaire de l'hérédité ?

Le **monde ARN** est une hypothèse résultant de l'idée que l'événement le plus critique se rapportant à l'origine de la vie repose sur l'émergence d'une molécule autoréplicative, une molécule qui puisse à la fois se copier elle-même et muter c'est-à-dire évoluer pour accroître l'efficacité de sa propre réplication. L'évolution fonctionne en effet sur la base de la variation et de la sélection et cette dernière se mesure toujours sous le sceau de la multiplication la plus efficace, la capacité à faire la meilleure entité en question possible.

Dans le monde d'aujourd'hui, l'ADN est le premier matériel génétique. Il est traduit grâce à un dispositif faisant appel à l'ARN et à un mécanisme de synthèse protéique spécifiant la position de 20 aminoacides dans les protéines enzymes. Trois systèmes moléculaires donc qui créent le paradoxe de ne pas livrer la clé du commencement. Ce paradoxe est cependant en passe d'être résolu aujourd'hui à la suite de deux découvertes capitales. La première est que l'ARN est une molécule biochimiquement plus primitive que l'ADN. La seconde est que l'ARN est capable d'effectuer le transfert de la liaison phosphodiester nécessaire à sa propre synthèse et possède ainsi une capacité enzymatique à l'instar des protéines.

En appui de la première découverte, il faut noter aussi que les désoxyribonucléotides sont obligatoirement synthétisés à partir de ribonucléotides précurseurs, que la synthèse de l'ADN nécessite des amorces ARN et que lorsqu'il s'agit de terminer la synthèse de l'ADN chromosomal à l'extrémité 3' de la chaîne ADN c'est une télomérase qui fonctionne en copiant une matrice ARN (voir le Mini Manuel de *Biologie moléculaire*).

L'hypothèse d'un monde initial où ce serait les protéines qui dirigeraient le transfert de l'information est très fortement improbable. Aucune forme d'interaction entre protéines n'est en effet connue, comme pouvant être apte à se copier, se répliquer et muter comme l'ARN. L'aspect le plus décisif de l'évolution moléculaire primitive réside dans la capacité des molécules à croître c'est-à-dire à s'autorépliquer dans des conditions compatibles avec la variation et la mutation, à fournir des nouveautés sur lesquelles la sélection naturelle pourra opérer afin d'améliorer la réplication dans un environnement changeant. C'est l'ARN qui réunit le mieux et de loin les propriétés nécessaires à l'émergence des molécules à partir desquelles se construiront les premiers systèmes. Ces derniers doivent avoir la double propriété d'être dépositaires d'une information génétique et être aptes à la mettre

en œuvre à la fois pour la transmettre par autoreplication, l'améliorer par mutation et l'exprimer en diverses fonctions par ses propriétés de catalyse.

La question capitale concernant le monde ARN est le classique paradoxe de la taille nécessaire pour fonctionner comme une enzyme. Une enzyme ARN devrait avoir une taille de 300 à 600 nucléotides pour, par la création de sites actifs, être réellement apte à réaliser diverses fonctions catalytiques. Or, il paraît peu vraisemblable qu'une telle création résulte d'un jeu aléatoire de 4^{300} possibilités, soit 10^{180} molécules. Walter Gilbert propose pour surmonter cette difficulté une solution qui recueille l'adhésion d'un nombre grandissant de biologistes moléculaires : les premières molécules d'ARN avaient une structure intron-exon.

Origine des introns et des exons, structure des gènes et évolution

Selon cette théorie, le matériel génétique ARN serait formé de courts modules, composés chacun de 30 à 40 bases (des exons ARN en quelque sorte), réunis par des introns **ARN autoépissables**. Cette structure linéaire serait bien adaptée à la formation d'une copie. Après l'opération, les introns s'autoépisseraient pour laisser ensemble une série d'exons qui se replieraient de telle manière à former un ribozyme. La présence d'introns résout trois difficultés et tout d'abord, celle du paradoxe de la taille de l'ARN. En effet, par le jeu d'un assemblage d'exons de petite taille (30 à 40 bases) une molécule à fonctions complexes peut en résulter. En second lieu, un moyen est fourni permettant de distinguer le matériel qui doit être copié pour produire des descendants, du matériel qui après épissage, se replie pour donner un ARN à fonction enzymatique. Enfin un matériel est obtenu qui en favorisant les recombinaisons illégitimes, stimule l'obtention de mutations sélectionnables.

L'aspect particulièrement novateur de ce modèle réside dans la formation d'une structure intron-exon apte à réaliser une transposition d'exons. En effet, dans l'hypothèse d'un autoépissage embarquant les deux introns enserrant un exon *(Fig. 7.1)* un transposon se forme qui peut se recombiner avec un autre intron présent au sein d'un gène ARN primitif accepteur. Un brassage d'exons en résulte, processus considéré aujourd'hui comme essentiel dans la formation des nouveaux gènes au cours de l'évolution. L'origine de ce type d'événement se situerait ainsi dès les premiers mécanismes moléculaires ayant conduit à l'émergence des premiers génomes dans un monde ARN.

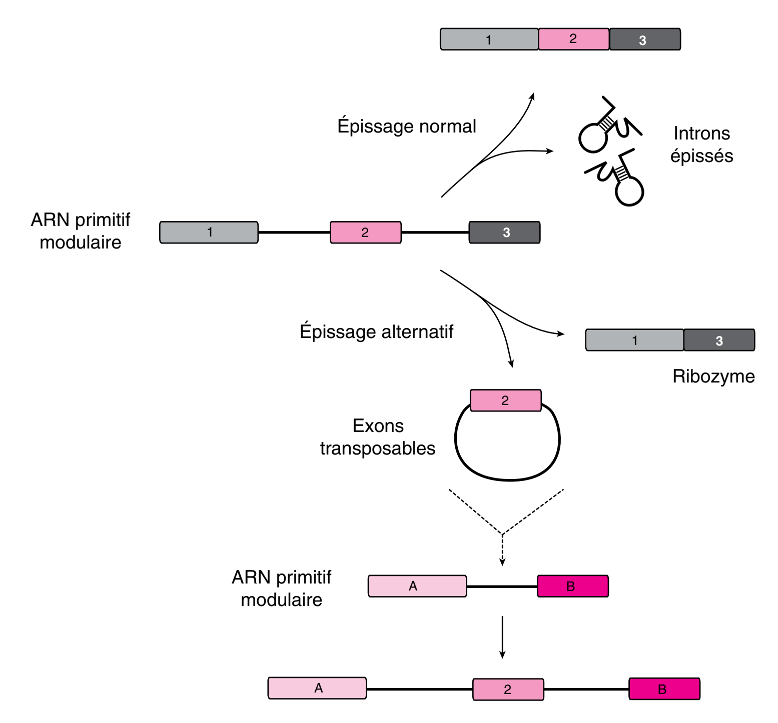


Figure 7.1 Représentation schématique d'une structure intron-exon au niveau d'un ARN primitif autoépissable.

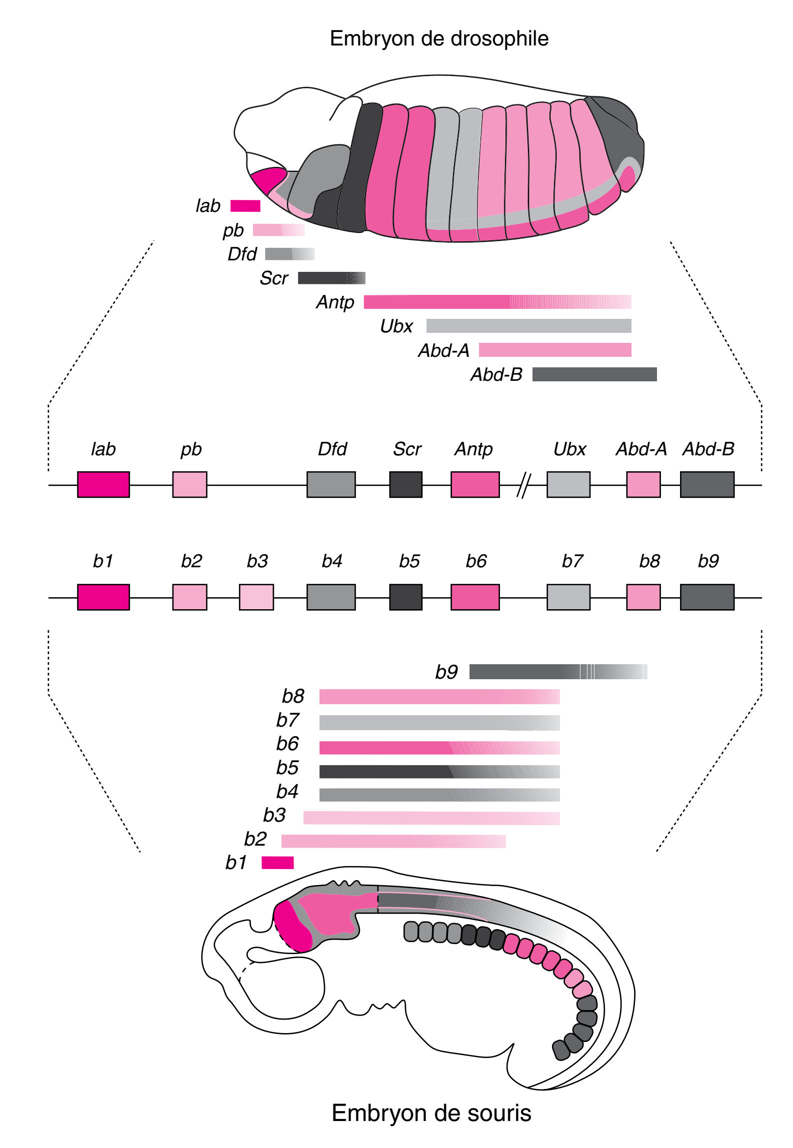
Alternativement : (i) les deux introns sont épissés pour produire un ribozyme (ii) les extrémités éloignées des deux introns épissent ensemble et produisent un exon d'intervention qui pourra occuper une nouvelle position au sein d'un nouveau gène.

7.3 LA BOÎTE À OUTILS GÉNÉTIQUES DU DÉVELOPPEMENT

Ce sont les différents allèles des gènes dits du développement qui dirigent le plan architectural du corps, le nombre, l'identité et la forme des différentes parties des organismes multicellulaires. Ces gènes sont hautement conservés parmi les grands phylums animaux. Ils ne représentent cependant qu'une petite fraction des génomes.

Figure 7.2 Organisation et expression des gènes *HOX* chez la drosophile et la souris.

Les huit gènes homéotiques de la souris sont organisés en deux clusters (complexes de gènes). Cinq des huit gènes sont localisés dans le complexe *Antennapaedia* et les trois autres dans le complexe *Bithorax*. Il y a une correspondance colinéaire entre l'ordre des gènes homéotiques le long du chromosome et les profils d'expression selon l'axe antéro-postérieur des embryons. Ainsi le gène *lab*, en position la plus extrême en 3′ du complexe *Antennapaedia* s'exprime dans la partie la plus ➤



➤ antérieure de la tête de l'embryon. Au contraire, le gène *Abd-B*, en position la plus 5' du complexe *Bithorax* est exprimé dans la région la plus postérieure. La signification de cette colinéarité n'est pas établie mais elle observée pour tous les embryons y compris chez les mammifères et l'homme. La souris possède 38 gènes *Hox* organisés en quatre clusters (un seul est présenté ici, le cluster *Hoxb*). Chaque cluster contient 9 ou 10 gènes homéotiques qui correspondent à chaque gène du cluster de la drosophile, avec le même type de colinéarité. Comme la souris, l'homme possède quatre clusters de gènes Hox avec les mêmes caractéristiques : *HOXA* (Chr 7), *Hoxb* (Chr 17), *HoxC* (Chr 12), *HOXD* (Chr 2).



Ils participent à des voies de signalisation, codent des facteurs transcriptionnels, des protéines de l'adhésion cellulaire, des récepteurs de la surface cellulaire et des morphogènes. Leurs fonctions sont fortement corrélées à leurs profils d'expression dans le temps et dans l'espace. L'évo-dévo se propose d'identifier tous ces gènes en s'attachant à définir leurs identités, produits, fonctions et interactions. À titre d'exemple, les gènes *HOX (Fig. 7.2)* , qui appartiennent à l'une des familles les plus importantes codent des facteurs de transcription dont la séquence comporte des homéoboîtes. Il s'agit de l'un des motifs de liaison aux protéines les plus largement distribués parmi les gènes du développement. Ils participent à la formation des axes de symétries antéro-postérieure et dorso-ventrale du corps. En spécifiant l'identité des différentes régions du corps, les gènes *HOX* déterminent là où les membres et autres segments du corps doivent se former chez l'embryon ou la larve. Des mutations dans l'un de ces gènes conduisent à la formation d'appendices sur-numéraires non fonctionnels chez les invertébrés. C'est le cas notamment de mutations dans le complexe *aristapedia* chez la drosophile. Ces mutations provoquent sur la tête, la formation d'une patte à la place d'une antenne (mutation connue sous le nom d'Antennapaedia).

Les gènes à homéoboîtes ont été tout d'abord identifiés chez les animaux à symétrie bilatérale. Certains gènes de Cnidaires, un groupe animal très primitif, possèdent également des domaines à homéoboîtes, ce qui au-delà d'apporter le chaînon manquant dans l'évolution des êtres vivants, apporte une preuve supplémentaire sur la manière dont procède l'évolution pour faire du neuf à partir du vieux, tout en conservant ce qui « marche ».

Les gènes homéotiques des plantes – les gènes à boîtes MADS – ne sont pas homologues des gènes HOX animaux. Plantes et animaux ne partagent donc pas les mêmes gènes homéotiques, ce qui suggère que les gènes homéotiques sont apparus deux fois au cours de l'évolution.

La régulation des gènes *HOX* s'avère très complexe. Elle implique des interactions réciproques pour la plupart inhibitrices.

Les gènes du développement

Tous les animaux débutent leur vie sous la forme d'une cellule unique, et pour cette raison, l'embryon s'auto-organise au cours du développement pour donner naissance aux nombreux types cellulaires différents, trouvés chez l'adulte. L'auto-organisation implique une série d'interactions géniques qui assemblent les cellules d'une manière qui préfigure le plan du corps adulte. Des événements de remodelage redistribuent les cellules en tissus spécifiques, et l'expression de batteries de gènes indispensables aux fonctions cellulaires spécialisées dirige ensuite les processus de différenciation. Le façonnage embryonnaire et la

différenciation finale sont intégrés par de vastes réseaux de régulation génique qui contrôlent par exemple les divisions asymétriques ou la signalisation entre cellules, grâce à des combinaisons de facteurs de transcription qui stimulent ou répriment leurs gènes cibles en se liant aux séquences *cis*-régulatrices. Si les gènes cibles codent eux-mêmes des facteurs de transcription, ils contrôleront alors d'autres gènes cibles si bien que le développement peut se comparer à un vaste réseau transcriptionnel où l'information *cis*-régulatrice serait la composante clé.

La notion d'homéoboîte

Une **homéoboîte** est une courte séquence d'ADN localisée au sein d'un gène impliqué dans la régulation du développement des animaux, champignons et plantes. Ces gènes sont nommés, pour cette raison, « gènes à homéoboîtes ». Ils forment la famille des gènes homéotiques.

La séquence ADN d'une homéoboîte possède environ 180 paires de bases et code un domaine protéique (l'**homéodomaine**) qui se lie à d'autres séquences d'ADN, d'où le nom générique de protéines de liaison à l'ADN. Les gènes homéotiques codent en général des facteurs de transcription qui activent à leur tour la transcription d'autres gènes qui participent aux cascades nécessaires à la formation d'une patte, par exemple. La protéine contenant une homéoboîte se fixe à sa cible ADN d'une manière spécifique. Habituellement cela ne suffit pas cependant pour qu'une protéine possédant une seule homéoboîte reconnaisse spécifiquement son gène cible. Les protéines à homéodomaines agissent en effet généralement sous la forme de complexes avec d'autres facteurs de transcription, eux-mêmes contenant des homéoboîtes, pour se lier à la région promotrice de leurs gènes cibles. Ces complexes ont des spécificités beaucoup plus élevées qu'une seule protéine à homéodomaine.

La notion de morphogène

Cette notion ancienne ne désigne pas une catégorie particulière de biomolécules, mais plutôt une substance qui contrôle la manière dont se développe un tissu et comment se positionnent les différentes cellules spécialisées qui le composent. Un **morphogène** diffuse généralement à partir d'une cellule émettrice. Il forme un gradient de concentration au sein du tissu cible en développement. Cette notion a été popularisée, il y a une cinquantaine d'années, par Lewis Wolpert qui a décrit comment un morphogène subdivisait un tissu en domaines cibles de l'expression de différents gènes, un peu à la manière des couleurs d'un drapeau. Le premier morphogène à être identifié fut un facteur de transcription dont le gradient, dans l'embryon syncitial de drosophile, participe à la définition de l'axe antéro-postérieur de



l'animal (voir plus loin). Un autre morphogène, homologue du TGFβ, nommé chez la drosophile Decapentaplegic (Dpp), fut ensuite identifié. Il agit aux stades tardifs du développement embryonnaire de la mouche.

Les **gradients de morphogènes** engendrent, au cours du développement de la mouche, différents types cellulaires. Ils façonnent l'organisation spatiale de l'embryon. Le morphogène subdivise, par l'effet d'un gradient de concentration, un champ de cellules à qui il fournit de ce fait une information spatiale qui induit ou maintient les mécanismes moléculaires de l'expression de différents gènes cibles sensibles à différents seuils de concentration (*Fig. 7.3*).

Figure 7.3 Un cluster de cellules libère une molécule de signalisation, le morphogène, qui diffuse vers les cellules cibles à travers la matrice extracellulaire.

Selon leur position plus ou moins distante du massif cellulaire émetteur, les cellules 1, 2 et 3 reçoivent des quantités différentes de morphogène. Le nombre de récepteurs de surface qui seront activés varie avec la distance et par voie de conséquence, les cellules contiendront de moins en moins de protéines régulatrices (facteurs de transcription par exemple) dans leur noyau. Une corrélation linéaire peut ainsi s'établir entre la quantité de morphogènes liés par les récepteurs et le niveau de transcription des gènes cibles.



Les cellules, les plus éloignées de la source de morphogène, recevront une faible concentration de morphogène et exprimeront les seuls gènes sensibles à ce faible niveau. Au contraire, les cellules les plus proches de la source seront au contact de fortes concentrations de morphogène et exprimeront les gènes sensibles à toutes les concentrations faibles et fortes. Différentes cellules seront ainsi conduites à se différencier, en réponse à des combinaisons de morphogènes de concentrations très variées. Le champ initial de cellules, se subdivisera en différents types cellulaires, suivant la position des cellules vis-à-vis de la source de morphogène. Il s'agit là d'un mécanisme général, responsable de la diversité des types cellulaires engendrés au cours du développement d'un animal. Les morphogènes les mieux connus sont des facteurs de transcription, mais la plupart sont des protéines sécrétées qui activent des voies de signalisation croisées entre les cellules cibles.

Le modèle historique de la drosophile

Tous les animaux actuels sont organisés suivant des axes de symétries antéro-postérieure et dorso-ventrale qui témoignent en quelque sorte

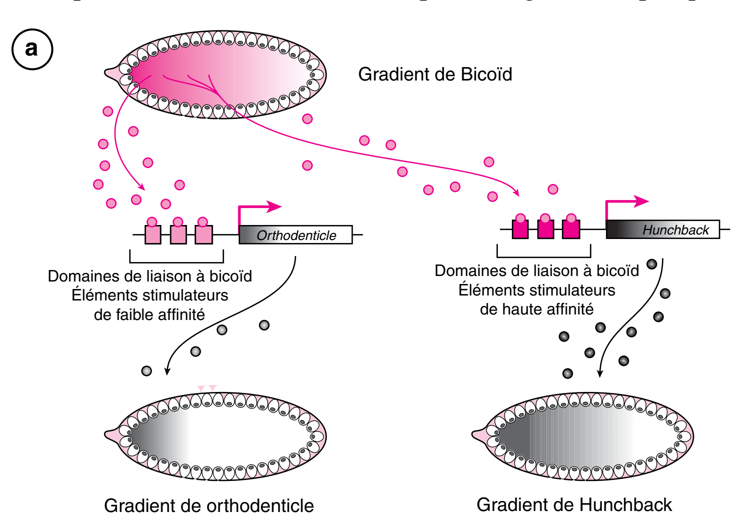
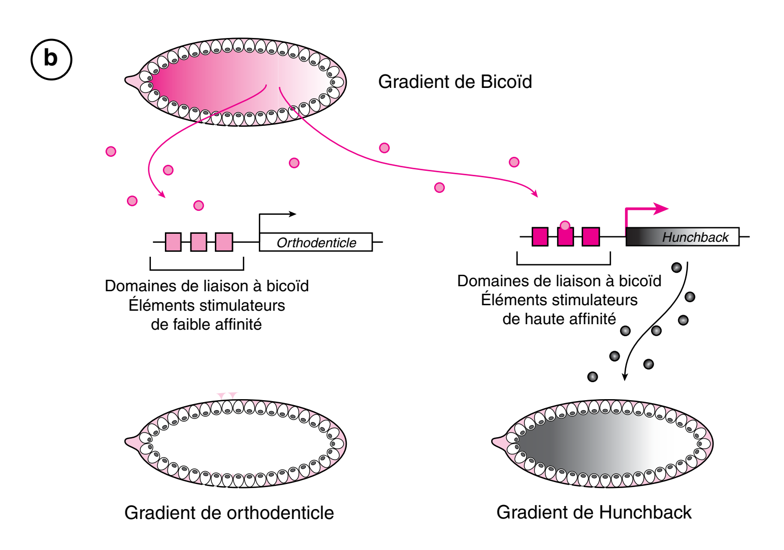


Figure 7.4 Exemple de l'intervention d'un gène à effet maternel, le gène *Bicoid*, dans la formation de l'axe de symétrie antéro-postérieure de la mouche.

a. Le gradient de la protéine bicoid, un facteur de transcription, engendre différents niveaux d'expression des gènes cibles, *orthodenticle* et *hunchback*. Seules de fortes concentrations de bicoid stimulent *orthodenticle* dans la tête de l'embryon. *Hunchback* est stimulé par des concentrations élevées et moyennes de bicoid, dans la tête et le thorax. Les deux gènes possèdent des sites ADN régulateurs en 5′ (« enhancers ») qui fixent la protéine bicoid et stimulent la transcription du gène en aval. Mais les trois sites « enhancers » de chacun des gènes n'ont pas la même affinité pour bicoid. Les sites de *orthodenticle* sont de faible affinité alors que les sites de *hunchback* sont à haute affinité.





b. Dans les parties centrales de l'embryon, le gène *orthodenticle* n'est pas stimulé car la concentration de la protéine bicoid n'est pas suffisante pour qu'une liaison avec les sites en 5′ de faible affinité se produise. Pour *hunchback* c'est le contraire, car la concentration de bicoid est suffisante pour que la protéine se lie avec les sites en 5′ de haute affinité du gène. Les protéines orthodenticle et hunchback sont aussi des facteurs de transcription qui contrôlent avec des effets de gradients (indiqués en noir grisé) d'autres gènes du développement qui participent à leur tour par leurs produits au façonnage de l'embryon de la mouche.

de la réussite complète au cours de l'évolution d'un type d'organisation basé sur un plan de symétrie bilatérale et une segmentation polarisée (*Fig. 7.4*, *Fig. 7.5* et *7.6*) dont le façonnage est contrôlé par des gènes très particuliers. Ainsi, les produits des gènes à effet maternel, ARNm ou protéines, sont synthétisés dans les tissus maternels entourant l'ovule (cellules folliculaires) puis importés dans ce dernier où, en association avec les produits des gènes embryonnaires homologues, leur distribution spatiale façonnera le plan du futur embryon et de l'adulte, en activant les réseaux de gènes appropriés. Le décryptage des principaux mécanismes génétiques à l'origine du façonnage précoce de l'embryon de la drosophile (mouche du vinaigre) constitue l'une des plus belles découvertes de la génétique du développement.

Le développement de *Drosophila melanogaster* commence par une série de treize divisions nucléaires qui ne sont pas accompagnées par la formation de cellules. L'embryon reste en effet une cellule unique avec 8 000 noyaux disposés près de la membrane jusqu'à la 14e division. À ce stade les membranes apparaissent et isolent les noyaux en cellules distinctes. Les facteurs de transcription, comme bicoid ou hunchback,

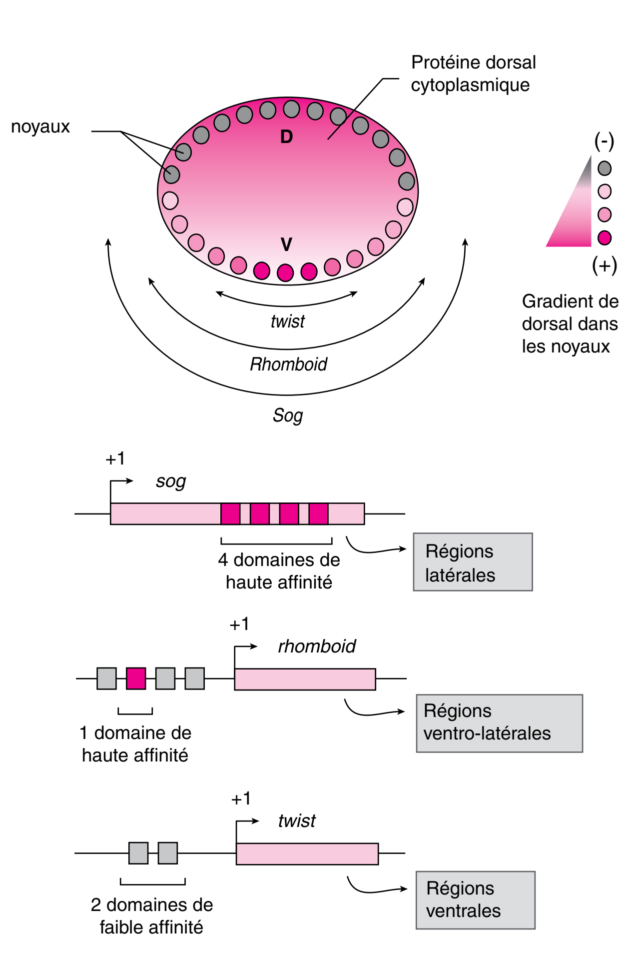


Figure 7.5 Exemple de l'intervention des trois gènes *sog*, *rhomboid* et *twist*, dans la formation de l'axe de symétrie dorso-ventrale de l'embryon de mouche.

Dans le cytoplasme syncitial (nombreux noyaux en l'absence de cellules) de l'embryon, la protéine Dorsal, un facteur de transcription, est fortement concentrée en partie dorsale et faiblement en partie ventrale. Par contre dans les noyaux c'est l'inverse, la protéine Dorsal est concentrée dans les noyaux de la partie ventrale et faiblement représentée dans ceux de la partie dorsale. Les deux sites régulateurs en 5′ du gène *twist* ont une faible affinité pour Dorsal et ne sont occupés que pour les fortes concentrations de Dorsal dans les noyaux de la partie ventrale. Un seul des sites en « cluster » 5′ du gène *rhomboid* possède une haute affinité pour Dorsal de sorte que ce gène n'est stimulé que dans les parties ventrale et latérale de l'embryon où la protéine Dorsal est assez bien représentée dans les noyaux. Les quatre sites introniques à haute affinité du gène *sog* peuvent lier la protéine Dorsal à toutes les concentrations nucléaires, ce qui a pour conséquence que ce gène est stimulé dans toutes les régions latérales de l'embryon où Dorsal est présente.

diffusent entre les noyaux et agissent comme des morphogènes, en produisant des gradients de concentration, sans le secours de voies de signalisation spécialisées qui sont habituellement nécessaires pour obtenir ce type d'effet. Dans ce dernier cas, les protéines s'attachent aux domaines extracellulaires des récepteurs transmembranaires et transfèrent, par un signal approprié, la concentration du morphogène au noyau (*Fig. 7.4* et *Fig. 7.5*).

Un lien direct s'établit alors avec les facteurs de transcription qui contrôlent l'expression des gènes. Les morphogènes sécrétés engendrent des gradients de concentration auxquels répondent les gènes cibles

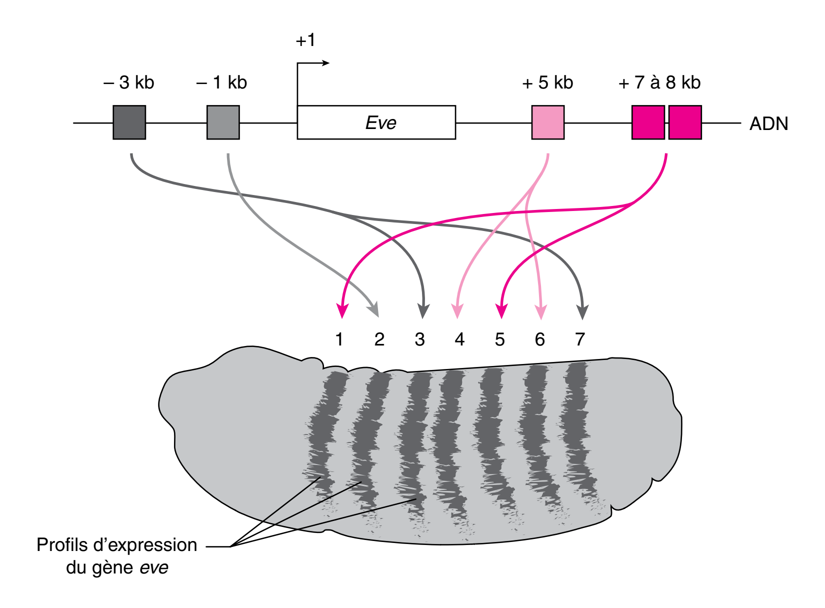


Figure 7.6 Exemple du contrôle de l'expression du gène *eve* participant à la segmentation du corps de l'embryon de drosophile.

La séquence codante de *eve* est assez courte, mais le gène lui-même possède des régions régulatrices de grande dimension (12 kpb), composées de sites localisés à la fois en 5′ et en 3′ de la région codante. La région 5′ possède deux stimulateurs (« enhancers »), d'une taille de 500 pb chacun, qui contrôlent l'expression du gène dans différents groupes de cellules disposées en forme de bandes (les bandes 2, 3 et 7) à la surface de l'embryon. La région régulatrice en 3′ possède trois stimulateurs qui contrôlent l'expression du gène *eve* dans les bandes de cellules 1, 4 et 6. Au total, les cinq stimulateurs produisent sept bandes d'expression cellulaire du gène *eve* qui avec d'autres gènes contribuent à la formation des sept parties segmentées du corps de l'embryon de drosophile.

selon le niveau de concentration auxquels ils sont sensibles. Le processus dépend du contrôle opéré par les motifs ADN régulateurs qui sont disposés généralement en amont 5' (parfois en 3' également) des gènes cibles et qui fixent avec des degrés d'affinité variable les facteurs de transcription régulateurs. Ces mécanismes généraux aboutissent à la différenciation de territoires cellulaires distincts *(Fig. 7.6)*. Ces derniers sont eux-mêmes à l'origine des tissus, organes et membres du futur organisme adulte. Cependant, de nombreux mécanismes additionnels restent encore à découvrir dans le vaste champ de la biologie et de la génétique du développement.

7.4 D'OÙ VIENT LA NOUVEAUTÉ EN MATIÈRE DE DÉVELOPPEMENT ?

Parmi toutes les données relatives à la génétique du développement, l'une des plus surprenantes se rapporte à la diversité phénotypique (morphologique) des organismes appartenant à de nombreux phyllums, diversité qui ne se reflète pas exactement dans la diversité de séquences géniques qui restent très souvent, malgré les distances phylogénétiques, étroitement apparentées, y compris pour des gènes qui ne sont pas directement liés au développement. Dès la formation des premiers génomes dans un monde ARN (voir plus haut), les mécanismes à l'origine des changements de séquences reposent sur des transpositions et des recombinaisons d'exons, réalisant un véritable brassage d'un petit nombre de modules seulement. D'autres changements, basés sur des mécanismes de duplications et de mutations ponctuelles, élargissent le champ des possibles, mais toujours à partir d'un nombre finalement assez restreint de modules génétiques dont les assemblages au cours de l'évolution ressemblent à un complexe jeu de meccano. À partir de quelques pièces de base, des constructions aux fonctions très diversifiées se réalisent dans une sorte de bricolage géant et patient, à l'échelle de millions d'années, soumis à une sélection aléatoire, et dont les traces sont déposées sous la forme de fossiles dans les couches géologiques.

Entre le génome humain (environ 25 000 gènes) et celui du nématode (19 000 gènes), un petit ver rond d'un millimètre de long, composé de 954 cellules, la génomique comparée a établi l'existence d'homologies et de similitudes de séquences codantes avoisinant les 36 % ! Avec la souris et certains autres mammifères, les homologies atteignent 80 %, et avec le chimpanzé, près de 99 %.

D'où provient alors la nouveauté, puisqu'elle ne se constate pas dans les génomes, au final si remarquablement apparentés ?

Ne résiderait-elle pas dans des changements, intéressant la régulation de l'expression des gènes et non dans les gènes eux-mêmes ?



L'évolution de l'expression génique

Une grande part de la biodiversité ne serait donc pas la conséquence de différences, pourtant bien réelles, présentes dans les gènes homologues des diverses espèces, mais plus concrètement de changements évolutifs intervenus dans la régulation de ces mêmes gènes. Ce type de changement, découvert récemment, induit en effet de fortes conséquences sur le fonctionnement des réseaux de gènes (voir chapitre 5) et beaucoup moins dans celui des gènes pris individuellement. L'exemple des gènes *HOX* (voir *Fig. 7.2*) montre – au-delà de différences phénotypiques entre les vertébrés et les invertébrés, différences liées à la duplication des gènes et à la formation de « clusters » chez les premiers – que ce sont des modifications dans la régulation de leur expression, à la fois dans le temps et dans l'espace, qui sont les plus pertinentes pour comprendre l'évolution des mécanismes génétiques entre les organismes considérés. Les changements intervenant dans les régulateurs opérant en *cis* notamment, sont beaucoup plus pertinents, que les changements dans le nombre de gènes ou la fonction des protéines. Ces changements de niveau, de profil et de rythme dans l'expression des gènes, fournissent finalement plus d'opportunités à la sélection naturelle que les seules mutations dans les produits des gènes eux-mêmes.

La régulation en *cis* et en *trans* de l'expression génique

La régulation en *cis* désigne la régulation qui dépend des séquences ADN régulatrices de l'expression d'un gène, quelles que soient leurs localisations. La régulation en *trans* désigne tous les facteurs protéiques ou ARN qui exercent leur fonction régulatrice en se liant notamment aux éléments de séquences en *cis*.

Les changements par mutation qui ont des effets régulateurs marqués sur l'expression génique au cours de l'évolution s'observent aux niveaux des séquences stimulatrices (enhancers), des facteurs de transcription, des séquences promotrices, et de divers autres éléments impliqués ou non dans la transcription. Les stimulateurs (enhancers) sont souvent cités comme étant les composants les plus pertinents. Une de leurs caractéristiques clés réside en effet dans leur aptitude à modifier de façon spécifique l'activité d'un ou de plusieurs gènes. De plus, leur structure modulaire *(Fig. 7.4*, *7.5* et *7.6)* et la nature combinatoire de leur mode opératoire, sont des caractéristiques tout à fait adaptées à la genèse de la diversité fonctionnelle. Cependant ces propriétés ne sont pas uniquement celles des stimulateurs. D'autres sites régulateurs, non transcriptionnels, présentent également des caractéristiques similaires. C'est le cas, par exemple, des sites localisés dans les 3' UTR. Ils régulent la traductibilité des transcrits. C'est le cas aussi des sites introniques et exoniques qui contrôlent les mécanismes d'épissage alternatif et



sont à l'origine de la diversité des transcrits. C'est le cas enfin des sites de liaisons pour les microARN régulateurs qui interviennent dans la stabilité des ARNm et leur traductibilité *(Fig. 7.7)*.

Nombreux sont les sites de régulation qui affectent les fonctions des gènes du développement d'une manière indépendante des « enhancers ». Les gènes sont en effet transcrits par des machineries biochimiques complexes dont le recrutement dépend d'interactions entre la chromatine et des éléments modulaires qui contrôlent la transcription (enhancers et silencers). Ces interactions moléculaires peuvent être modifiées par d'autres éléments comme les isolateurs. Des éléments de séquences au niveau des promoteurs ainsi que des facteurs additionnels spécifiques de tissus participent au processus de démarrage de la transcription et

Différentes étapes intervenant dans l'expression d'un gène

(1) (2)
(4)
Intron
ADN
(3) Exon
(5) Transcription
pré ARNm
(6) (7)
(8)
Épissage
ARNm
(9)
(10) Export dans le cytoplasme
ARNm (11)
(12)
Traduction
(13)
Protéine (14)
Localisation de la protéine
Fonction de la protéine

Différents niveaux de régulation possibles

Chromatine
- Dynamique du noyau (1)
- condensation de la chromatine et remodelage - système de mémoire (2)

Promoteur :
- Spécificité du promoteur (3)
- Régulation par des facteurs de transcription (enhancers, inhibiteurs) (4)

Transcription
- Dégradation des facteurs de transcription (5)
- Régulation de la polyadénylation (6)

Maturation de l'ARN
- Régulation par les petits ARN (7)
- Dégradation de l'ARN (8)
- Epissage alternatif (9)
- Export de l'ARN dans le cytoplasme (10)

Traduction
- Localisation subcellulaire de l'ARNm (11)
- Dégradation de l'ARNm (12)
- Régulations traductionnelles (13)

Modifications post-traductionnelles (phosphorylations, ubiquitylation, sumoylation et glycosylation) (14)

Figure 7.7 Schéma récapitulatif des niveaux où s'exerce une régulation en *cis* et en *trans* de l'expression génique.



contribuent à la cinétique de ce démarrage. Au niveau de la transcription, les enhancers ne sont donc pas les seuls éléments qui influencent ce processus. Mais ce n'est que le début. Les transcrits ARN sont modifiés, au cours de leur transcription et la nature de ces événements (épissage, « capping », polyadénylation, épissage croisé, etc.) a une énorme influence sur la qualité et la quantité des ARN messagers matures. Ces derniers sont par ailleurs soumis à des systèmes de contrôle qualité/dégradation puis exportés du noyau selon un processus lui aussi objet de divers contrôles. Certains messages possèdent des séquences particulières qui définissent leur localisation subcellulaire où ils seront finalement traduits. Les modules localisés dans les UTRs affectent les cinétiques de traduction et la demi-vie de ces messages. Une fois la protéine synthétisée, d'autres processus régulateurs interviennent, comme la phosphorylation, la glycosylation, l'ubiquitinylation et la sumoylation. La vitesse et le rythme de ces événements régulateurs qui affectent la fonction de la protéine et sa demi-vie sont modulés eux-mêmes au cours du développement par des gènes spécifiques *(Fig. 7.7)*.

Estimation simplifiée de la taille moyenne des sites régulateurs mutables pour un gène standard

À la question de savoir : quelles sont les tailles respectives des sites transcriptionnels et non transcriptionnels dont les mutations peuvent avoir des effets sur la régulation de l'expression d'un gène et par voie de conséquence imposer des changements évolutifs majeurs si la sélection naturelle s'exerce positivement, un calcul simple permet de répondre.

Taille moyenne des sites régulateurs transcriptionnels

Les calculs sont basés sur l'hypothèse qu'un gène possède en moyenne sept modules stimulateurs indépendants, chacun d'eux contenant dix sites de liaison pour des facteurs de transcription. Chaque site de liaison d'un facteur de transcription se compose de six paires de bases, une paire seulement n'étant pas spécifique de la liaison. Avec ces données, le calcul donne :

- nombre de stimulateurs : 7 ;
- nombre de sites de liaison pour les facteurs de transcription par stimulateur : 10 ;
- taille d'un site : 5 pb (6 dont 1 non spécifique).

Soit une taille totale moyenne de 350 pb ($7 \times 10 \times 5$) pour la région stimulatrice mutable régulant l'expression d'un gène.

Taille moyenne des sites régulateurs non transcriptionnels

Les calculs pour estimer les dimensions des autres sites mutables régulateurs de l'expression des gènes sont basés sur les longueurs

moyennes de la séquence minimale liée par l'ARN polymérase et du signal consensus de polyadénylation (AATAAA), sur la longueur et le nombre de sites d'épissage pour un gène moyen possédant quatre exons et trois introns. On admet la présence de cinq stimulateurs d'épissage par exon avec une longueur moyenne de six nucléotides soit la longueur moyenne des sites de liaisons de nombreuses protéines liant l'ARN. La longueur moyenne des modules régulateurs de 3' et 5' UTR sont conformes aux données de la littérature. Sur ces bases, le calcul donne :

- promoteur minimal (boîte TATA, initiateur, motif stimulateur de l'ARN polymérase) : 35 pb ;
- signal de polyadénylation : 6 pb ;
- sites accepteurs d'épissage : 3 pb × 2 = 6 pb ;
- sites donneurs d'épissage : 3 pb × 2 = 6 pb ;
- sites stimulateurs d'épissage : 6 pb × 20 = 120 pb ;

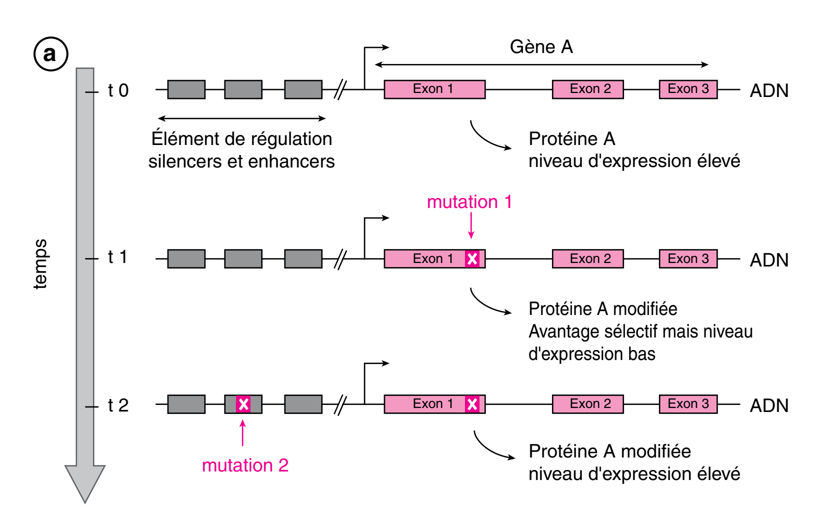


Figure 7.8 a Deux scénarios montrant comment des mutations de sites régulateurs de l'expression de gènes sont sélectionnables au cours de l'évolution.

Une première mutation (mutation 1) intervient dans un site exonique d'épissage en 3' (exon 1 du gène A). La nouvelle protéine A présente une séquence peptidique modifiée comparée à la protéine A de type sauvage, mais le phénotype qu'elle influence s'en trouve favorisé par la sélection naturelle. Pourtant le niveau d'expression de la nouvelle protéine n'est pas optimal et une pression de sélection s'exercera pour qu'une deuxième mutation au temps t_2 se produise dans un autre site régulateur ou dans un facteur protéique, transcriptionnel ou non. Ici, l'exemple retenu est celui d'une mutation 2 dans un site régulateur transcriptionnel en *cis* dont la conséquence sera de stimuler la transcription du gène *A* muté au temps t_1.

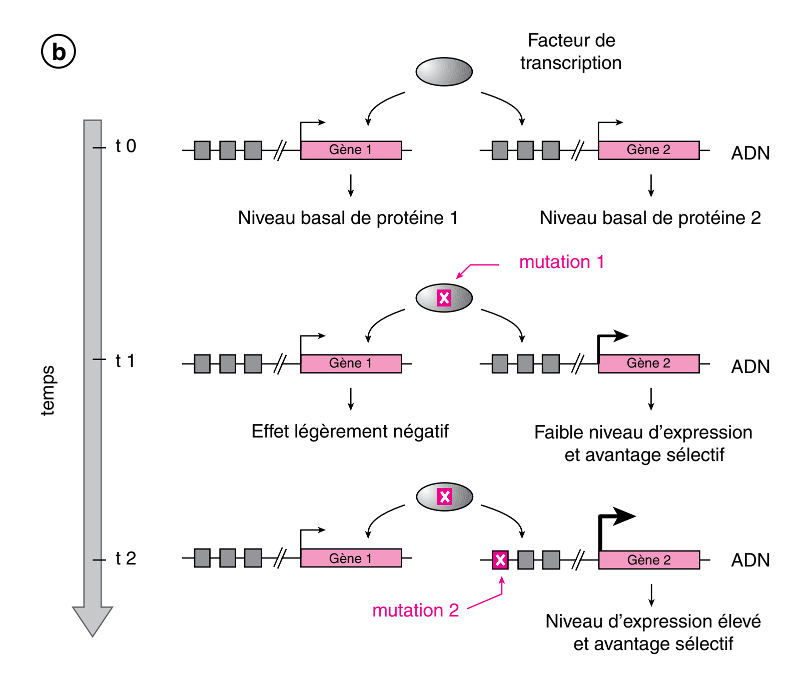


Figure 7.8 b Deux scénarios montrant comment des mutations de sites régulateurs de l'expression de gènes sont sélectionnables au cours de l'évolution.

Une première mutation (mutation 1) intervient au temps t_1 dans un facteur agissant en *trans*. Cela peut être un facteur de transcription comme retenu ici, mais aussi un facteur agissant par exemple sur la stabilité du messager du gène 2, avec pour conséquence d'exercer un effet légèrement négatif sur le gène 1, et au contraire un effet légèrement positif sur le gène 2, pas complètement décisif cependant en terme d'avantage sélectif. Une deuxième mutation (mutation 2) au temps t_2 dans un site régulateur stimulateur de la transcription du gène 2 contribuera à rendre optimal le niveau de transcription de ce gène et l'avantage sélectif qui l'accompagne. On observera ici que les deux mutations ne sont probablement pas liées. La première mutation peut donc être perdue lors d'un événement de recombinaison ou par ségrégation, ce qui éliminera son effet négatif sur le gène 1.

- modules régulateurs en 5' (IRES) : 30 pb ;
- modules régulateurs en 3' (HSP 83, β-actine) = 247 pb.

Soit une taille moyenne totale de 450 pb (35 + 18 + 120 + 30 + 247) pour les sites mutables non transcriptionnels régulant l'expression d'un gène.

Ces estimations montrent que les tailles respectives des sites régulateurs mutables sont comparables qu'ils soient transcriptionnels ou non transcriptionnels. Elles montrent finalement que, comparativement aux séquences codantes, l'impact des mutations des sites régulateurs de l'expression des gènes peut jouer un rôle majeur dans les changements évolutifs à l'origine de la biodiversité des processus développementaux *(Fig. 7.8)*. On notera que les signaux de phosphorylation,



d'ubiquitinylation et de sumoylation, de rétention nucléaire, d'adressage aux organites, les signaux peptidiques et autres signaux régulateurs n'ont pas été pris en compte.

Modifications épigénétiques de l'expression génique

Les modifications épigénétiques de la régulation de l'expression des gènes et de la formation des phénotypes accompagnées secondairement par des mutations géniques constituent une autre classe de mécanismes générateurs de nouveautés évolutives en matière de développement (voir chapitre 5, et aussi le Mini Manuel de *Biologie moléculaire*). Certains biologistes considèrent qu'en favorisant l'explosion des formes et des plans d'organisation caractéristiques d'une macroévolution intense, ce type de changements a pu jouer un rôle majeur au début de l'histoire évolutive des organismes multicellulaires.



POINTS CLEFS

- L'évolution s'appuie sur le constat que les seuls individus qui survivent (qui sont sélectionnés) sont ceux qui, dans un environnement changeant, sont les plus aptes à se reproduire parce qu'ils y sont les mieux adaptés. L'environnement exerce donc un tri, une sélection qui provoque une évolution en faisant apparaître des adaptations.
- On désigne par le terme de macroévolution celle qui intervient au niveau de l'espèce ou à un niveau supérieur. Typiquement la macroévolution fait référence aux changements qui modifient l'embryogenèse, les plans d'organisation, les principales caractéristiques des individus et qui sont à la base de la phylogénie des grands groupes du monde vivant. La microévolution par contre désigne au sein d'une population des changements évolutifs de moindre ampleur comme ceux qui sont relatifs à la fréquence des allèles.
- Les biologistes considèrent que macroévolution et microévolution sont liées par le fait que l'accumulation graduelle de petits changements entraîne les changements de plus grande ampleur, même si des sauts évolutifs rapides sont également visibles dans les traces paléontologiques.
- L'évolution du développement biologique, ou en abrégé l'évo-dévo, se propose de comparer les processus développementaux des organismes vivants (animaux et des plantes). Elle essaie de déterminer les relations ancestrales qui existent entre les organismes et comment ces processus ont évolué.



- L'évo-dévo aborde l'origine et l'évolution du développement embryonnaire. Elle analyse comment les changements dans le développement et les processus qui le dirigent ont produit de nouvelles caractéristiques biologiques. Elle s'intéresse au rôle de la plasticité développementale dans l'évolution et comment l'écologie intervient dans ces phénomènes.
- Le monde ARN est une hypothèse résultant de l'idée que l'événement le plus critique se rapportant à l'origine de la vie repose sur l'émergence d'une molécule autoréplicative, une molécule qui puisse à la fois se copier elle-même et muter c'est-à-dire évoluer pour accroître l'efficacité de sa propre réplication. L'évolution fonctionne en effet sur la base de la variation et de la sélection et cette dernière se mesure toujours sous le sceau de la multiplication la plus efficace, la capacité à faire la meilleure entité en question possible.
- Tous les animaux débutent leur vie sous la forme d'une cellule unique, et pour cette raison, l'embryon s'auto-organise au cours du développement pour donner naissance aux nombreux types cellulaires différents, trouvés chez l'adulte. L'auto-organisation implique une série d'interactions géniques qui assemblent les cellules d'une manière qui préfigure le plan du corps adulte. Des événements de remodelage redistribuent les cellules en tissus spécifiques, et l'expression de batteries de gènes indispensables aux fonctions cellulaires spécialisées dirige ensuite les processus de différenciation.
- Le façonnage embryonnaire et la différenciation finale sont intégrés par de vastes réseaux de régulation génique qui contrôlent par exemple les divisions asymétriques ou la signalisation entre cellules, grâce à des combinaisons de facteurs de transcription qui stimulent ou répriment leurs gènes cibles en se liant aux séquences *cis*-régulatrices. Si les gènes cibles codent eux-mêmes des facteurs de transcription, ils contrôleront alors d'autres gènes cibles si bien que le développement peut se comparer à un vaste réseau transcriptionnel où l'information *cis*-régulatrice serait la composante clé.
- Les changements par mutation ont des effets marqués sur l'expression des gènes au cours de l'évolution. Ils s'observent à plusieurs niveaux : séquences stimulatrices (enhancers), facteurs de transcription, séquences promotrices, et divers autres éléments impliqués ou non dans la transcription. Les modifications épigénétiques de la régulation de l'expression des gènes et de la formation des phénotypes constituent une autre classe de mécanismes générateurs de nouveautés évolutives en matière de développement.



QCM - QROC

7.1 Le terme « évolution » signifie-t-il progrès et amélioration ? Expliquer.

7.2 Qu'est-ce que l'évo-dévo ?

7.3 Citer quatre mécanismes qui contribuent à l'évolution des gènes.

7.4 Qu'est-ce qu'un morphogène ?

7.5 Dans l'embryon de Drosophile, des séquences situées dans les 3' UTR sont responsables de la localisation des ARNm des gènes *bicoid* et *oskar* respectivement aux pôles antérieur et postérieur de l'embryon. Quel phénotype peut-on s'attendre à observer si à la suite de diverses constructions et opérations de transferts moléculaires, les régions 3' UTR sont échangées et qu'une femelle homozygote pour les deux gènes mutants est obtenue ?

7.6 Plusieurs gènes, dits à effets maternels, interviennent dans la formation des axes de symétrie antéro-postérieure et dorso-ventrale. Des femelles mutantes homozygotes pour certains de ces gènes produisent des embryons avec des défauts de segmentation quel que soit le génotype des derniers, mutant hétérozygote ou homozygote (exemple du gène *bicoid* : *bcd+/bcd* ou *bcd/bcd*). Pour d'autres gènes mutés, à effets maternels, la production d'un phénotype mutant peut être évitée si un allèle sauvage du gène considéré est apporté par le père. Comment peut-on expliquer la différence exercée sur le phénotype selon l'une ou l'autre des catégories de gènes à effets maternels.

7.7 Les supports physiques des mécanismes régulateurs intervenant dans les relations de type « hiérarchie de gènes » et « réseaux de gènes » sont-ils différents ? Expliquer.

7.8 Donner un exemple de site régulateur transcriptionnel et non transcriptionnel.

RÉPONSES

7.1 Non. Les notions de supériorité d'un génotype, d'un phénotype ou d'un organisme sur un autre sont étrangères à la théorie de l'évolution. Évolution ne signifie pas amélioration. Les notions clés sont les notions de variation, d'adaptation et de sélection au hasard.

7.2 L'évo-dévo essaie de déterminer les relations ancestrales qui existent entre les organismes et de quelle manière ces processus ont évolué. L'évo-dévo aborde l'origine et l'évolution du développement embryonnaire. Elle analyse de quelle manière les changements dans le développement et les processus qui le dirigent ont produit de nouvelles caractéristiques biologiques. Elle s'intéresse au rôle de la plasticité développementale dans l'évolution et de quelle façon l'écologie (l'environnement) intervient dans ces phénomènes.

7.3 La mutation, le brassage, la transposition et la duplication.

7.4 Un morphogène est une biomolécule qui intervient directement par son activité, souvent exprimée sous la forme d'un gradient spatio-temporel, dans la morphogenèse d'un organisme.

7.5 Une inversion de l'axe antéro-postérieur.

7.6 Les gènes dits à effet maternels sont localisés dans les cellules nourricières de l'embryon et non dans l'embryon lui-même qui ne reçoit que les ARNm ou les protéines correspondantes. Si le gène est muté, l'ARNm ne sera pas fonctionnel par défaut de localisation (voir 7.5) ou par inactivité de la protéine correspondante. La présence d'un gène nucléaire sauvage homologue, apporté par le père évitera le phénotype mutant résultant pourvu que l'expression de ce gène s'effectue correctement à la fois dans l'espace (localisation du transcrit ou de la protéine dans la partie attendue de l'embryon) et dans le temps (expression au moment adéquat de l'embryogenèse). Les phénotypes liés aux gènes à effets maternels mutés qui n'ont pas d'homologue nucléaire fonctionnel ne seront pas évités.

7.7 Non. Dans la majorité des cas, le fonctionnement des cascades régulatrices ou des réseaux régulateurs repose sur des éléments *cis*- (motifs de séquences ADN) et *trans*-régulateurs (facteurs protéiques et ARN régulateurs) communs ou similaires. Les différences sont dans les modalités de mise en œuvre (existence ou non de gradients, nombre de sites *cis*-régulateurs, affinité différentielle des sites, et surtout interactions croisées ou non des éléments *cis* et *trans*-régulateurs).

7.8 Les sites régulateurs transcriptionnels du gène *eve* et les sites régulateurs en 3' UTR des transcrits de *bicoid.*



Génétique des populations

PLAN

OBJECTIFS

- ➤ Aborder les premiers éléments de génétique des populations
- ➤ Aborder les calculs de fréquences alléliques et génotypiques
- ➤ Aborder les bases du modèle de Hardy-Weinberg

En publiant en 1859 *De l'origine des espèces*, Darwin exposait une théorie de l'évolution des espèces par sélection naturelle qui peut se résumer en trois principes :

- ➤ le **principe de variation** : dans une population, il existe une variation entre individus et elle concerne leur morphologie, leur physiologie et leur comportement ;
- ➤ le **principe de l'hérédité** : les jeunes ressemblent plus à leurs géniteurs qu'à des individus auxquels ils ne sont pas apparentés ;
- ➤ le **principe de sélection** : certains individus sont plus aptes à survivre et à se reproduire dans un environnement donné.

Il en ressort que les populations naturelles se composent d'individus génétiquement différents pour un grand nombre de gènes et qu'il existe un **polymorphisme génétique** des populations. Confrontées aux modifications des facteurs environnementaux, ce polymorphisme constitue pour les espèces un véritable **capital adaptatif** permettant leur survie.

La génétique des populations s'efforce d'évaluer la diversité génétique et d'établir des lois décrivant dans le temps et l'espace le maintien ou la modulation de la variation génétique au sein d'une population. Son point de départ est la recherche des différents allèles d'un gène dans une population concernée et **l'analyse des fréquences** de ceux-ci. Pour un grand nombre de gènes, l'analyse du polymorphisme donnera

une estimation de la diversité génétique globale rencontrée entre des individus d'une même population mais aussi entre des populations *(Tableau 8.1)*. Cependant, les modèles théoriques de la génétique des populations (les lois) ne s'appliquent correctement aux mécanismes responsables de la diversité génétique que pour un voire deux gènes, rarement plus.

Tableau 8.1 Fréquences des groupes sanguins MN dans diverses populations.

Populations	Fréquences phénotypiques (ou génotypiques*)			Fréquences des allèles	
	M/M	M/N	N/N	p (M)	q (N)
Indiens américains	0,60	0,35	0,05	0,78	0,22
Blancs américains	0,29	0,50	0,21	0,54	0,46
Noirs américains	0,28	0,50	0,22	0,53	0,47
Esquimaux	0,83	0,16	0,01	0,91	0,09
Allemands	0,30	0,50	0,20	0,55	0,45
Chinois	0,33	0,49	0,18	0,57	0,43
Nigériens	0,30	0,50	0,20	0,55	0,45

* Fréquences phénotypiques et fréquences génotypiques sont ici identiques parce qu'à chacun des génotypes correspond un phénotype parfaitement identifiable.

Le système de groupe sanguin MN résulte de variations ponctuelles dans la séquence en acides aminés de la partie N-terminale de la glycophorine A, protéine présente à la surface des érythrocytes. Les phénotypes sont mis en évidence à l'aide d'anticorps dirigés contre le produit de l'un ou l'autre des deux allèles. L'antigénicité est déterminée par le polymorphisme des acides aminés 1 et 5. Pour l'allèle M la séquence N-terminale de la glycoprotéine est : **Ser**-Ser-Thr-Thr-**Gly**-Val-etc., et pour l'allèle N : **Leu**-Ser-Thr-Thr-**Glu**-Val-etc. Pour une population donnée, la somme des fréquences (phénotypiques ou alléliques) est bien évidemment égale à 1.

8.1 CALCULS DES FRÉQUENCES GÉNOTYPIQUES ET ALLÉLIQUES

Pour un locus donné, la génétique des populations cherche à comprendre la relation entre deux variables étroitement liées : la fréquence génotypique et la fréquence allélique ou génique.

$f_{(A)}$ se dit fréquence de l'allèle *A* ou fréquence allélique ; $f_{(G)x}$ se dit fréquence génotypique pour x qui représente ici la combinaison des allèles *A/A*.



Les fréquences génotypiques se calculent par le comptage des individus porteurs des différents génotypes possibles (voir par exemple *Tableau 8.1*). Ainsi, pour un locus où deux allèles *A* et *A'* sont identifiés, on pourra dénombrer les individus diploïdes porteurs de l'un des trois génotypes *A/A, A/A'* et *A'/A'*. Les fréquences génotypiques attribuables à chaque groupe d'individus seront par exemple $f_{(G)x}$, $f_{(G)y}$ et $f_{(G)z}$ respectivement. Généralement, ces fréquences sont exprimées soit en pourcentage, leur somme étant alors égale à 100 ; soit en proportion, la somme égalant alors 1.

Connaissant les fréquences génotypiques on peut calculer les fréquences alléliques. Exemple : Pour calculer la fréquence de l'allèle *A*, on doit s'appuyer sur les fréquences génotypiques des individus porteurs de l'allèle *A*. Or la fréquence de l'allèle *A* correspond à la probabilité de trouver l'allèle *A* chez un individu pris au hasard dans la population considérée.

Pour déterminer une fréquence génotypique, on considère ici qu'à chacun des génotypes homozygotes (*A/A* et *A'/A'*) ainsi qu'au génotype hétérozygote (*A/A'*) correspond un phénotype parfaitement identifiable.

Si l'individu examiné possède un génotype homozygote *A/A*, la probabilité de trouver l'allèle *A* sera donc maximale et égale à 1. Évidemment, cela n'est vrai que pour ce type d'individus. Aussi, faut-il tenir compte de leur représentation dans la population, autrement dit de la fréquence x du génotype dans la population, soit $f_{(G)x}$. Pour cette partie de la population, on dira que la fréquence allélique $f(A)$ est égale à $1 \times f_{(G)x}$.

Si l'individu examiné possède un génotype *A/A'*, la probabilité de trouver l'allèle *A* sera évidemment de 1 sur 2, soit 0,5, qu'il faudra corriger par la fréquence $f_{(G)y}$ correspondant aux individus de génotype *A/A'* dans la population. Pour cette partie de la population, on dira que la fréquence allélique $f_{(A)}$ est de $0{,}5 \times f_{(G)y}$ ou $\frac{f_{(G)y}}{2}$.

Enfin, si l'individu examiné possède un génotype *A'/A'*, la probabilité de trouver *A* sera nulle, corrigée en théorie par la fréquence $f_{(G)\,z}$ des individus de génotype *A'/A'*, soit une fréquence de l'allèle *A* pour cette partie de la population égale à $f_{(A)} = 0 \times f_{(G)z}$.

Pour l'ensemble de la population, la fréquence $f_{(A)}$ ou p sera égale à la somme des fréquences partielles, soit :

$$f_{(A)} = (1 \times f_{(G)\,x}) + (0{,}5 \times f_{(G)y}) + (0 \times f_{(G)z}) = f_{(G)x} + \frac{f_{(G)y}}{2} = p\ ;$$



Un raisonnement similaire amène pour la fréquence $f_{(A')}$ ou q de l'allèle *A'* à :

$$f_{(A')} = (0 \times f_{(G)x}) + (0{,}5 \times f_{(G)y}) + (1 \times f_{(G)z}) = \frac{f_{(G)y}}{2} + f_{(G)z} = q\ ;$$

la somme des fréquences $f_{(A)}$ et $f_{(A')}$ étant bien évidemment égale à 1, d'où :

$$f_{(A)} + f_{(A')} = p + q = 1.$$

8.2 LE MODÈLE DE HARDY-WEINBERG

Dans une population **d'individus diploïdes**, différents génotypes peuvent coexister, c'est le polymorphisme génétique. Pour un gène donné, les fréquences de ces génotypes $f_{(G)}$ évoluent au fil des générations. Quand on peut les établir à une génération donnée, est-il possible de les connaître à la génération suivante ? C'est à cette question que s'efforce de répondre le modèle mathématique ou loi de Hardy-Weinberg.

Conditions requises pour l'application du modèle

Le modèle ou loi de Hardy-Weinberg s'applique à une population théorique formant une communauté d'individus qui ne se perpétuent entre eux que par la reproduction sexuée.

La loi théorique est établie dans le cas le plus simple en considérant un gène autosomique n'ayant que deux allèles possibles. Ceci exclut l'existence de toute mutation qui ne ferait qu'accroître le nombre des allèles. De plus *(Fig. 8.1)* nous devons considérer que :

- la formation des couples reproducteurs (1) s'opère au hasard, c'est-à-dire indépendamment des liens de parenté et des génotypes. Ces couples sont dits **panmictiques** (on parle également de conditions de **panmixie**). Il n'y a pas de sélection au niveau des reproducteurs.
- les deux types de gamètes quel que soit leur génotype (leur allèle respectif) ont la même possibilité de participer à la fécondation (2).
- les zygotes (3), quelle que soit la combinaison allélique (génotype) parmi les trois possibles, sont supposés avoir la même probabilité de donner un individu (4). Pour le gène considéré, aucun des trois génotypes n'influence l'espérance de vie (5) et tous les individus adultes à la génération suivante (n + 1) peuvent se reproduire.
- enfin, la taille de la population doit être suffisamment grande pour négliger les fluctuations aléatoires de la fréquence des allèles.



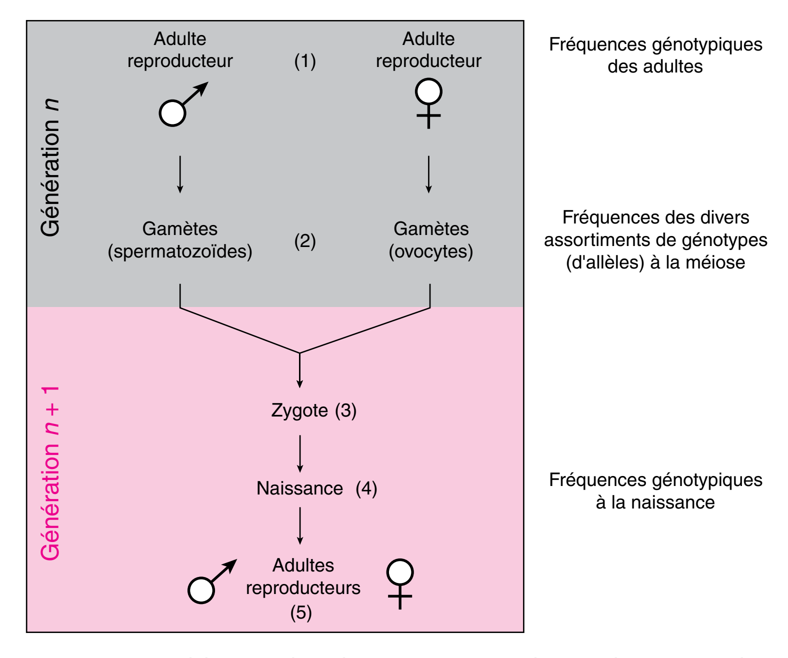


Figure 8.1 Modèle général, appliqué en génétique des populations, entre les stades adultes reproducteurs de deux générations successives.

Les chiffres entre parenthèses indiquent les points critiques restrictifs qui sont détaillés dans le texte.

Généralisation du modèle de Hardy-Weinberg

On considère une population constituée d'adultes reproducteurs, à la génération n. Pour un gène autosomique diallélique (allèle *A* et allèle *A'*) trois génotypes possibles sont détectables : *A/A*, *A/A'* ou *A'/A* et *A'/A'* avec des fréquences respectives $f_{(G)x}$, $f_{(G)y}$ et $f_{(G)z}$. Si la fréquence de l'allèle *A*, $f_{(A)}$ ou p, est identique dans les deux types de gamètes (spermatozoïdes et ovocytes), il en est de même pour celle de l'allèle *A'*, $f_{(A')}$ ou q, la somme des féquences des deux types gamétiques (p + q) sera égale à 1. Les unions au hasard des gamètes lors de la fécondation dans une telle population panmictique (voir plus haut) peuvent se représenter comme dans la *figure 8.2*.

La probabilité que le gamète mâle (spermatozoïde) et le gamète femelle (ovocyte) soient tous deux porteurs de l'allèle *A* sera égale au produit (p × p) des fréquences alléliques, soit p^2. Cette valeur définit également la fréquence génotypique $f_{(G)\,x}$ des homozygotes *A/A* à la génération suivante (n + 1). De même, la fréquence $f_{(G)z}$ des homozygotes *A'/A'* sera $q \times q = q^2$ et celle des hétérozygotes *A/A'*(génotypes *A'/A* et *A/A'*) $f_{(G)y}$ sera $(p \times q) + (q \times p) = 2\,pq$.

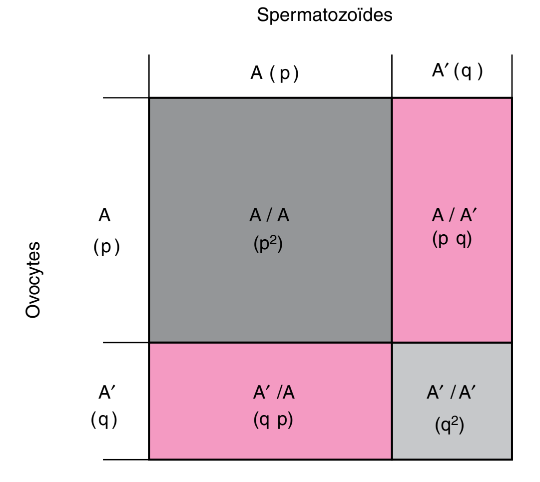


Figure 8.2 Union au hasard des gamètes lors de la fécondation dans une population panmictique.

Ainsi, à la génération (n + 1), après un croisement au hasard, les trois génotypes *A/A*, *A/A'*, *A'/A'* se répartissent respectivement selon les fréquences génotypiques $f_{(G)x} = p^2$, $f_{(G)y} = 2\ pq$, $f_{(G)z} = q^2$.

Si on applique la formule générale de calcul des fréquences alléliques à partir des fréquences génotypiques, présentée plus haut, on obtient à la génération (n + 1) pour :

➤ la fréquence de l'allèle *A* :

$$f_{(A)} = f_{(G)x} + \frac{f_{(G)y}}{2} = p^2 + \frac{2\ pq}{2} = p\ (p + q)$$

➤ la fréquence de l'allèle *A'*:

$$f_{(A')} = f_{(G)z} + \frac{f_{(G)y}}{2} = q^2 + \frac{2\ pq}{2} = q\ (p + q)$$

Or comme p + q = 1, on retrouve pour les deux allèles *A* et *A'* les fréquences p et q identiques à celles établies à la génération précédente n. Cette stabilité de la fréquence des allèles est appelée « équilibre de Hardy-Weinberg ». Elle requiert pour être maintenue indéfiniment les conditions suivantes : une population panmictique, de taille infinie, sans mutation pour les allèles considérés, et dans laquelle ne s'exerce aucune sélection. Un dernier critère pour le maintien de l'équilibre est l'absence de migration.

Si les fréquences alléliques $f_{(A)} = p$ et $f_{(A')} = q$ demeurent stables, les fréquences génotypiques $f_{(G)x}$, $f_{(G)y}$ et $f_{(G)z}$ qui avaient à la génération n des valeurs quelconques, déterminées expérimentalement, sont devenues à la génération (*n* + 1) respectivement p^2, $2\ pq$ et q^2 (voir



plus haut). Elles demeureront inchangées tant que les fréquences alléliques p et q ne seront pas soumises à des variations. Cet état d'équilibre entre les fréquences alléliques et génotypiques d'un gène autosomique est atteint, dans ces conditions restrictives un peu théoriques, en une seule génération.

Mise en évidence de la relation entre équilibre allélique et équilibre génotypique

Considérons deux populations P1 et P2 où ne sont représentés que les 2 allèles *A* et *A'*. Ces deux populations seront caractérisées à titre d'exemple, par les fréquences génotypiques suivantes *(Tableau 8.2)*.

TABLEAU 8.2 FRÉQUENCE GÉNOTYPIQUES DE DEUX POPULATIONS P1 ET P2.

Génotype Fréquence	**A/A** $f_{(G)x}$	**A/A'** $f_{(G)y}$	**A'/A'** $f_{(G)z}$
P1	0,3	0,2	0,5
P2	0,1	0,6	0,3

Les fréquences parentales des allèles *A* et *A'* dans ces deux populations s'avèrent identiques et égales à 0,4 et 0,6 respectivement :

➤ Population 1 :

$$f_{(A)} = f_{(G)x} + \frac{f_{(G)y}}{2} = 0,3 + \frac{0,2}{2} = 0,4$$

$$\text{et } f_{(A')} = f(G)z + \frac{f_{(G)y}}{2} = 0,5 + \frac{0,2}{2} = 0,6$$

➤ Population 2 :

$$f_{(A)} = f_{(G)x} + \frac{f_{(G)y}}{2} = 0,1 + \frac{0,6}{2} = 0,4$$

$$\text{et } f_{(A')} = f_{(G)z} + \frac{f_{(G)y}}{2} = 0,3 + \frac{0,6}{2} = 0,6$$

Ainsi, bien que les fréquences génotypiques diffèrent (voir *tableau*), ces deux populations possèdent les mêmes fréquences alléliques : p = 0,4 et q = 0,6. Après une génération de panmixie, soit à (*n* + 1) les deux populations posséderont cependant les mêmes fréquences génotypiques :

$$f_{(G)x} = p^2 = (0,4)^2 = 0,16$$

$$f_{(G)y} = 2\,pq = 2\,(0,4 \times 0,6) = 0,48$$

$$f_{(G)z} = q^2 = (0,6)^2 = 0,36$$

qui se maintiendront tant que p et q resteront constants.



8.3 APPLICATIONS AUX ALLÈLES RARES

Jusqu'ici nous avons considéré qu'il était possible de déterminer les fréquences génotypiques d'après l'observation du phénotype des individus. Ceci signifie que la combinaison des deux allèles du gène considéré présent chez l'hétérozygote donne un phénotype aisément identifiable distinct des phénotypes des deux homozygotes. Nous avons donc fait abstraction des caractères dominants ou récessifs qui rendent impossible le calcul des fréquences alléliques car, par exemple pour les deux génotypes (*A/A* et *A/a*) nous observerions un phénotype identique lié au caractère dominant (*A*). Le modèle de Hardy-Weinberg permet cependant d'avoir accès indirectement aux fréquences alléliques si la population obéit bien à l'équilibre précédemment décrit. Dans ces conditions, les fréquences génotypiques $f_{(G)x}$, $f_{(G)y}$ et $f_{(G)z}$ pour lesquelles on ne peut estimer, par l'observation des deux phénotypes, que les valeurs $[f_{(G)x} + f_{(G)y}]$ et $f_{(G)z}$, se présentent à la génération $(n + 1)$ comme suit *(Tableau 8.3)* :

TABLEAU 8.3

	Phénotype dominant		**Phénotype récessif**
Génotype	A/A	A/a	a/a
Fréquence génotypique	$f_{(G)\,x} = p^2$	$f_{(G)\,y} = 2\,pq$	$f_{(G)\,z} = q^2$

Le phénotype récessif ne peut provenir que d'individus homozygotes portant les deux allèles *a*. La fréquence d'apparition de tels individus (q^2) dans une population est quantifiable par simple comptage.

On en déduit la fréquence allélique de l'allèle *a*, $f_{(a)} = \sqrt{q^2}$, sachant que p + q = 1, on en déduit indirectement $f_{(A)} = p = 1 - q$. À partir de la connaissance d'une des fréquences alléliques (q), il s'avère donc possible de représenter la relation générale entre les fréquences des différents génotypes, homozygotes et hétérozygotes *(Fig. 8.3)*. Dans une population panmictique, on observe qu'un allèle rare, c'est-à-dire dont la fréquence est proche de 0, ne se maintiendra que grâce à sa présence chez les individus hétérozygotes. Cette observation possède un grand intérêt dans le cas de maladies génétiques récessives pour déterminer dans une population le nombre de porteurs sains présents (voir chapitre 1).

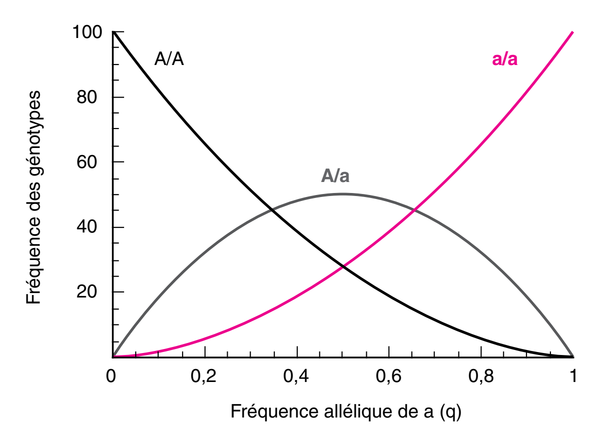


Figure 8.3 Représentation des fréquences génotypiques observables dans une population à l'équilibre en fonction de la fréquence allélique (q) de l'allèle récessif *a*.

Pour reprendre deux exemples déjà décrits, l'albinisme et la phénylcétonurie, on observe les fréquences suivantes dans la population française *(Tableau 8.4)* :

TABLEAU 8.4

	Fréquence de la maladie à la naissance*	Fréquence allélique $\sqrt{q^2}$	Fréquence des porteurs sains 2 pq ou 2 q(1 – q)
Albinisme	1/10 000	1/100	1,98/100
Phénylcétonurie	1/16 000	0,8/100	1,59/100

* Fréquence du phénotype ou génotype récessif q^2

Le rapport individus porteurs sains/individus malades est égal à :

$$\frac{2\,pq}{q^2} \text{ soit } \frac{2q(1-q)}{q^2} \text{ ou } \frac{2(1-q)}{q}$$

Il montre que plus un allèle est rare dans une population, plus la proportion d'hétérozygotes (porteurs sains) est supérieure à celle des homozygotes (malades). Dans l'exemple ci-dessus, ce rapport est égal à 198 dans le cas de l'albinisme (fréquence allélique 1/100) et 254,4 pour la phénylcétonurie (fréquence allélique 0,8/100). Autrement dit, plus ce rapport est élevé, plus la fréquence de la mutation au locus responsable de la maladie est rare.



La même démarche peut être suivie pour les mutations responsables de maladies génétiques dominantes. Dans ce cas, la fréquence *F* des individus malades inclut les hétérozygotes et *F* a comme valeur : $p^2 + 2\,pq$. La fréquence de l'allèle sain est $q = \sqrt{1 - F}$ et celle de l'allèle muté, vecteur de la maladie, $p = 1 - \sqrt{1 - F}$.

Ces calculs supposent une absence de sélection ou d'interventions extérieures comme c'est le cas chez l'homme en raison des progrès de la médecine. Il n'en demeure pas moins que la loi d'équilibre de Hardy-Weinberg reste un modèle central à partir duquel il devient possible de suivre l'évolution d'une population (limite et vitesse des modifications) lorsque les conditions d'équilibre ne seront plus respectées. On peut voir ci-dessous le rôle de la consanguinité dans ce phénomène.

8.4 PARENTÉ ET COEFFICIENT DE CONSANGUINITÉ

Dans les populations naturelles, les accouplements aléatoires sont en général la règle, et la distribution des allèles à un locus donné reflète cette situation. Mais le phénomène n'est pas universel et les écarts à la panmixie peuvent par exemple provenir d'unions préférentielles entre individus apparentés. On parle d'**endogamie** ou **consanguinité**. En génétique, est dit consanguin le fils (la fille) issu(e) de l'union de deux individus génétiquement apparentés, c'est-à-dire ayant au moins un ancêtre commun.

Les termes : père, mère, fils, fille, etc., s'appliquent également en génétique animale pour désigner l'appartenance des animaux à telle ou telle génération.

Dans les populations naturelles, l'endogamie peut être la conséquence d'un confinement géographique, de certains mécanismes de reproductions ou de caractéristiques de comportement. Dans les populations fortement sélectionnées, l'endogamie est un risque permanent. C'est le cas dans les élevages d'animaux de ferme, et le sélectionneur doit veiller constamment à déjouer les conséquences néfastes des accouplements entre individus apparentés. Bien que chez l'homme la consanguinité soit un tabou social, les unions entre individus semblables (**homogamie**) pour certains caractères (taille, couleur de la peau, etc.) sont communes. Une différence doit être faite cependant entre l'homogamie qui ne concerne qu'un caractère particulier, même si quelques gènes sont impliqués, et l'endogamie qui concerne la totalité du génome.

Le degré de consanguinité peut s'exprimer par le **coefficient de parenté** (r). Il manifeste la probabilité qu'un ancêtre ait transmis à ses deux descendants directs le même allèle parmi les deux distincts qu'il

possède à un locus donné. Pour un frère et une soeur, il sera de 0,5. Cette valeur représente la proportion de génome identique chez deux individus apparentés, conséquence de leur ascendance commune. C'est pourquoi, ensuite l'accouplement de deux individus apparentés accroît la probabilité qu'un allèle provenant d'un ancêtre commun soit présent chez les descendants où il s'exprimera sous la forme homozygote. Ainsi, la consanguinité n'augmente pas la fréquence des mutations, mais elle permet en revanche une augmentation de la proportion des individus homozygotes. La probabilité d'homozygotie par filiation est appelée **coefficient de consanguinité** (F).

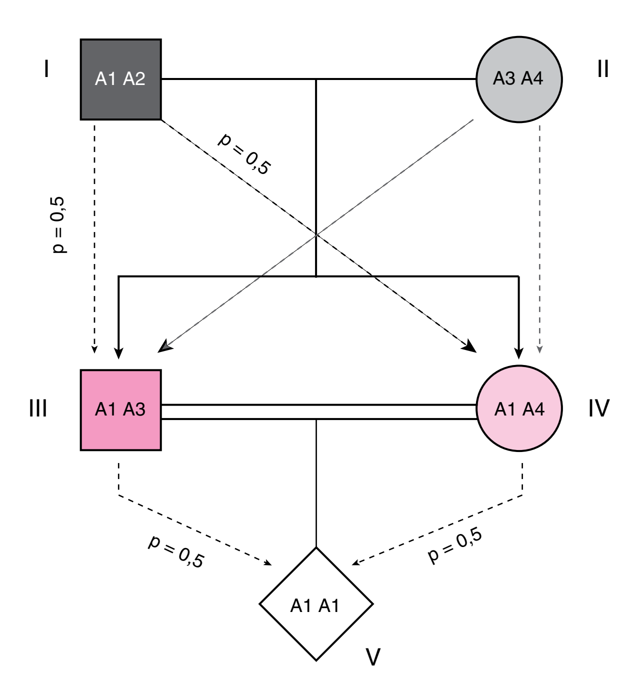


Figure 8.4 Le coefficient de consanguinité : probabilité d'homozygotie par filiation.

La *figure 8.4* matérialise cette probabilité. Elle est de 0,5 pour que chaque individu (III et IV) reçoive l'allèle *A1* de leur père (I). De même, elle est de 0,5 pour que cet allèle *A1* soit transmis à leur propre descendant (V) consanguin, soit une probabilité de 0,25. Cette valeur exprime également la probabilité que V soit homozygote par filiation. Une consanguinité aussi forte peut avoir des conséquences désavantageuses lorsque l'allèle est rare et qu'à l'état homozygote il provoque une maladie. Si sa fréquence p est de 1/1000, la fréquence d'apparition

d'homozygotes de l'allèle défectueux A_1 dans une population naturelle d'après la loi de Hardy-Weinberg sera de p^2 soit 1/1 000 000 (voir exemple de la phénylcétonurie au chapitre I). Dans le cas d'un couple d'individus apparentés (III) et (IV) dont l'un des parents (I) ou (II) est hétérozygote (porteur sain), le risque que ce couple ait une descendance (V) homozygote augmente. Mathématiquement la probabilité diminue, elle correspond à la probabilité que (I) ou (II) soit hétérozygote (fréquence de l'allèle de 2 pq), qu'ils transmettent l'allèle rare défectueux A_1 à leurs deux descendants (III) et (IV) soit une probabilité de 0,5 et que (V) soit homozygote (probabilité de 0,25), ce qui donne :

$$(2\,pq + 2\,pq) \times 0{,}5 \times 0{,}5 \times 0{,}25 = \frac{pq}{4}$$

Comme q (fréquence de l'allèle correct) est voisine de 1, la probabilité que le descendant (V) soit malade (homozygote pour l'allèle défectueux) sera de p/4 soit 1/4 000. Si on compare la fréquence d'un homozygote (p^2) dans une population naturelle panmictique, à celle obtenue suite à l'union consanguine entre individus directement apparentés (p/4), il en ressort une augmentation d'un facteur 250 soit $(1 \times 10^6)/(4 \times 10^3)$ de la probabilité d'homozygotie pour l'allèle défectueux, soit un risque accru identique que le descendant manifeste la maladie génétique.

De manière plus générale, dans le cas d'unions consanguines, on peut écrire que la fréquence d'homozygotes pour un allèle quelconque est de :

$$\frac{p/4}{p^2} = \frac{1}{4\,p}.$$

L'homozygotie par filiation décroît en fonction du degré de parenté. Par exemple elle est de $\frac{1}{16\,p}$ pour des cousins au premier degré.

En favorisant les unions consanguines chez les animaux domestiques ou par autofécondation chez les végétaux, c'est-à-dire en pratiquant **l'endogamie systématique**, l'éleveur peut, à partir d'une population hétérozygote, en quelques générations « créer » une population homogène pour un phénotype par la fixation d'allèles impliqués dans le déterminisme du caractère recherché. Cette démarche s'applique en sélection bovine où, par exemple l'un des caractères phénotypiques d'identification d'une « race » est la couleur de la robe.

8.5 MODÉLISATION DE LA SÉLECTION NATURELLE

Il faut se rappeler que la sélection opère à partir du phénotype des individus. C'est parce que tel individu est plus apte à survivre et peut par conséquent mieux se reproduire que son génotype se répandra au



sein de la population. Envisageons un modèle dans lequel la sélection naturelle des individus s'applique à un locus possédant deux allèles, *B* et *B'*. La probabilité de survie de l'individu, de la naissance jusqu'à l'âge adulte, varie en fonction du génotype. Considérons les probabilités de survie suivantes *(Tableau 8.5)* :

TABLEAU 8.5 PROBABILITÉ DE SURVIE EN FONCTION DES GÉNOTYPES.

Génotypes	Probabilité de survie
B/B	1
B/B′	1
B′/B′	1 – s

On appelle **coefficient de sélection**, s, une valeur comprise entre 0 et 1 qui traduit la différence entre la plus faible aptitude de survie des individus homozygotes (*B'/B'*)*'* jusqu'à l'âge adulte et celle des individus possédant les génotypes les plus aptes (*B/B*, *B/B'*). Par exemple, si s = 0,2, cela signifie que le probabilité de survie ou **aptitude à la survie** des individus de génotype *B'/B'* vaut 80 % de celle des individus *B/B* et *B/B'*. Au fil des générations, il y aura une dilution de l'allèle *B'* aboutissant à son élimination. On dira alors qu'il y a **fixation** de l'allèle *B*. En génétique des populations, on considère qu'un allèle est fixé lorsque sa fréquence est de 1, c'est-à-dire que tous les individus sont homozygotes pour cet allèle. La vitesse de variation de la fréquence des allèles se détermine en calculant la fréquence p' de *B* à une génération (n + 1) par rapport à la fréquence p du même allèle à la génération précédente.

$\Delta p = p' - p$ représente la variation de la fréquence allélique d'une génération à la suivante. La sélection s'opère à la génération (n + 1) et elle modifie le nombre d'adultes dans la mesure où les individus homozygotes (B'/B') adultes sont moins nombreux qu'à la naissance. Les nouvelles fréquences génotypiques sont donc à calculer de telle manière que leur total demeure égal à 1. Pour cela, il faut diviser chaque valeur de fréquence par la somme des fréquences relatives calculées ainsi : $p^2 + 2\,pq + q^2\,(1 - s) = 1 - sq^2$.

À la génération (n + 1), pour les individus d'âge adulte, la fréquence allélique p' de *B* sera égale à la somme de la fréquence génotypique des individus *B/B*, soit p^2, et de la moitié de celle des individus *B/B'*, soit 2 pq, 2, divisées par la somme des fréquences relatives $1 - sq^2$, d'où :



$$p' = \frac{p^2 + pq}{1 - sq^2}, \text{ avec } p + q = 1 \text{ soit } p' = \frac{p}{1 - sq^2} \quad (1)$$

On en déduit que

$$\Delta p = p' - p = \frac{spq^2}{1 - sq^2} \quad (2)$$

Le coefficient de sélection ayant une valeur toujours positive, $1 - sq^2$ est inférieur à 1, aussi p' devient à la génération (n + 1) supérieur à p. Grâce à l'équation (2), le calcul de la variation des fréquences alléliques s'avère possible de génération en génération connaissant le coefficient de sélection et les fréquences alléliques de départ. Si seule, la variation des fréquences alléliques est connue, on pourra calculer s en combinant les équations (1) et (2) d'où :

$$s = \frac{\Delta p}{p'q^2}$$

Si dans le cas traité, l'allèle B était un allèle dominant, sa fréquence augmenterait au fil des générations. Cependant, il faudrait un très

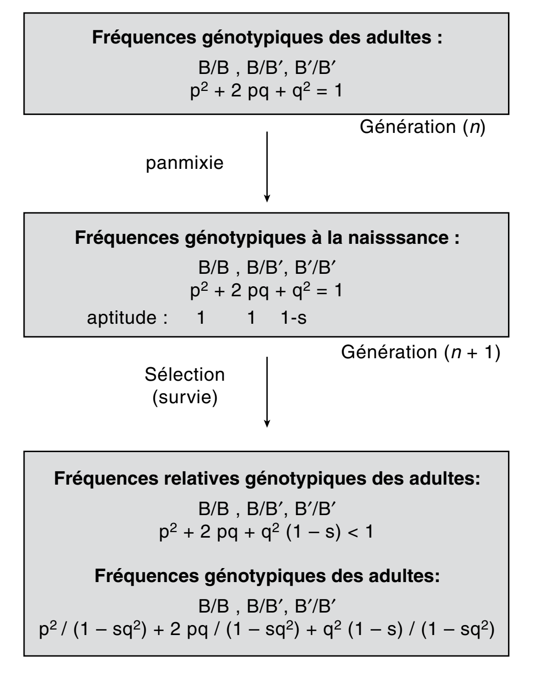


Figure 8.5 Schéma récapitulatif montrant la variation des fréquences génotypiques lors d'une sélection naturelle.



grand nombre de générations pour éliminer complètement l'allèle *B'* récessif. En effet, plus sa fréquence devient rare plus la proportion liée au génotype hétérozygote (*B/B'*) augmente *(Fig. 8.5* et *Tableau 8.6)*. Comme les phénotypes *B/B* et *B/B'* sont identiques, la sélection n'a aucune prise sur l'allèle *B'* contenu dans l'hétérozygote où il demeure « à l'abri ». Il ne peut donc pas être totalement éliminé.

La panmixie permet uniquement la conservation de ces fréquences entre deux générations *n* à l'âge adulte et (*n* + 1) à la naissance. À la génération (*n* + 1), le coefficient de sélection entraîne une modification des fréquences chez les adultes. Un exemple numérique *(Tableau 8.6)* illustre ceci avec une valeur hypothétique pour le coefficient de sélection s de 0,3 pour les homozygotes *B'/B'*.

TABLEAU 8.6 EXEMPLES DE VARIATIONS CHIFFRÉES LORS D'UNE SÉLECTION NATURELLE.

	Génotypes			Total
	B/B	***B/B'***	***B'/B'***	
À la naissance : Nombre	4	32	64	100
Fréquence (%)	0,04	0,32	0,64	1
Aptitude	1	1	0,7	
À l'âge adulte : Nombre	4	32	45	81
Fréquence relative	0,04	0,32	0,45	0,81
Fréquence (%)*	0,05	0,40	0,55	1

* On constate qu'en une génération, dans la population concernée, la proportion d'individus B/B et B/B' a augmenté en raison de la sélection qui s'exerce sur les individus B'/B' entre la naissance et l'âge adulte.



POINTS CLEFS

- Les populations naturelles se composent d'individus génétiquement différents pour un grand nombre de gènes et qu'il existe un polymorphisme génétique des populations. Confrontées aux modifications des facteurs environnementaux, ce polymorphisme constitue pour les espèces un véritable capital adaptatif permettant leur survie.
- La génétique des populations s'efforce d'évaluer la diversité génétique et d'établir des lois décrivant dans le temps et l'espace le maintien ou la

modulation de la variation génétique au sein d'une population. Son point de départ est la recherche des différents allèles d'un gène dans une population concernée et l'analyse des fréquences de ceux-ci.

- Dans une population d'individus diploïdes, différents génotypes peuvent coexister, c'est le polymorphisme génétique. Pour un gène donné, les fréquences de ces génotypes $f_{(G)}$ évoluent au fil des générations. Quand on peut les établir à une génération donnée, est-il possible de les connaître à la génération suivante ? C'est à cette question que s'efforce de répondre le modèle mathématique ou loi de Hardy-Weinberg.
- Dans les populations naturelles, les accouplements aléatoires sont en général la règle, et la distribution des allèles à un locus donné reflète cette situation. Mais le phénomène n'est pas universel et les écarts à la panmixie peuvent par exemple provenir d'unions préférentielles entre individus apparentés. On parle d'endogamie ou consanguinité. En génétique, est dit consanguin le fils (la fille) issu(e) de l'union de deux individus génétiquement apparentés, c'est-à-dire ayant au moins un ancêtre commun.

QCM - QROC

8.1 On observe que dans une population soumise au test de sensibilité au PTC, 70 % des individus sont sensibles. En supposant que ce phénotype est lié à un allèle dominant *S* et l'insensibilité au PTC à l'allèle récessif *s*, quelles sont les fréquences des allèles et des génotypes dans cette population

8.2 Dans une population le taux de mutation de $G \rightarrow g$ est de 4×10^{-6}. Si $p = 0,8$ peut-on prédire ce qu'il sera dans 50 000 générations ?

8.3 Une population essentiellement composée d'individus fertiles de génotypes *A/A* et *A/a* comporte 0,1 % d'individus *a/a*. Quelle pression de sélection s'exerce contre les individus de génotype a/a si le taux de mutation $A \rightarrow a$ est de 10^{-5} ?

RÉPONSES

8.1 Puisque 70 % des individus sont sensibles au PTC, 30 % sont insensibles. La fréquence des homozygotes récessifs est q^2, d'où $q = \sqrt{0,30} = 0,55$.

Comme $p + q = 1$ on a : $p = 1 - q = 1 - 0,55 = 0,45$.



D'où : $p^2 = 0{,}45^2 = 0{,}20$ (*S/S*) ; $2\,pq = 0 \times 0{,}45 \times 0{,}55 = 0{,}50$ (*S/s*)
$q^2 = 0{,}30$ (*s/s*)

8.2 La fréquence allélique devient de plus en plus faible au cours des générations. Δp également. Après 50 000 générations on a :

$$p = 0{,}8 \times e^{-(50\ 000 \times 4 \times 10^{-6})} = 0{,}65.$$

8.3 0,01



Glossaire

Pour se repérer plus facilement, les mots du glossaire sont en couleur dans le cours.

Adduct : divers groupes chimiques réactifs d'origine environnementale, accolés par voie enzymatique ou chimique à l'ADN et provoquant des erreurs de réplication.

ADN polymérase : famille d'enzymes qui synthétisent de nouveaux brins d'ADN à partir d'une matrice modèle.

Allèle : une des diverses formes d'un gène présente à un locus

ARN non codant : désigne tout ARN résultant de la transcription d'une séquence d'ADN jamais traduit en protéine.

Bactériophage : virus qui infecte une bactérie.

Centromère : région spécialisée de l'ADN de chaque chromosome eucaryote maintenant ensemble les chromatides sœurs. C'est aussi le site de liaisons des protéines du kinétochore qui capturent ensuite les microtubules à partir des fuseaux mitotiques et méiotiques.

Centrosome : organite organisateur de microtubules agissant comme pôle du fuseau pendant la mitose. Il contient souvent une paire de centrioles.

Chromatide : copie d'un chromosome par réplication de l'ADN jointe encore à l'autre copie par le centromère.

Chromatine : complexe macromoléculaire du noyau d'une cellule eucaryote formé par l'ADN, les histones, certains ARN et d'autres protéines non-histones. Matériau constitutif des chromosomes.

Chromosome : arrangement linéaire de gènes et d'autres formes d'ADN, associé parfois avec certains ARN et protéines.

Chiasma : structure en forme de croix observée entre des chromatides non-sœurs lors de la méiose.

Clone : population de cellules ou d'organismes génétiquement identiques, issus d'une même cellule ou d'un même ancêtre.



Chromosome polytène : chromosome géant produit par un processus endomitotique au cours duquel de nombreux segments d'ADN restent liés.

Cluster : région regroupant des gènes ayant des fonctions similaires mais d'une structure différente en modifiant la régulation de leur expression, à la fois dans le temps et dans l'espace.

Codominance : situation dans laquelle un hétérozygote présente de manière équivalente les effets phénotypiques des deux allèles.

Coiffe : désigne l'addition à l'extrémité 5' d'un ARN messager d'une méthylguanine phosphate, protégeant le dit ARN d'une dégradation exonucléasique.

Complémentation : production d'un phénotype de type sauvage par deux mutations différentes combinées au sein d'une cellule diploïde ou d'un hétérocaryon.

Conjugaison : union temporaire de deux organismes unicellulaires (bactéries, protistes,…) au cours de laquelle une cellule donneuse transfère du matériel génétique à une cellule receveuse.

Croisement test : croisement d'un individu de génotype inconnu ou d'un hétérozygote avec un individu de génotype double récessif pour le ou les caractères étudiés.

Crossing-over : échange de fragments chromosomiques semblables entre des chromatides non sœurs homologues.

Contigs : groupes ordonnés de clones chevauchants qui assemblés, reconstituent une région chromosomique ou un génome.

Cycle cellulaire : cycle de reproduction de la cellule oscillant entre la mitose (M) et l'interphase (G_1, S et G_2) et constitué d'une séquence d'événements par lesquels la cellule duplique son contenu avant de se diviser en deux cellules.

Cytogénétique : approche cytologique de la génétique par une étude en microscopie des chromosomes.

Diploïde : cellule ou organisme qui possède deux jeux complets de chromosomes homologues (c'est-à-dire deux allèles pour chaque locus génétique). Les cellules somatiques possèdent 2 n chromosomes.

Division équationnelle : division qui maintient constant le nombre de chromosomes ou niveau de ploïdie des cellules.

Division réductionnelle : division nucléaire qui génère des noyaux-fils possédant chacun la moitié des chromosomes contenus dans le noyau parental.

Domaine activateur : région d'une protéine qui interagit directement avec le domaine carboxy-terminal de l'ARN polymérase et stimule ainsi la transcription.

Dominance : désigne l'effet d'un allèle ou d'un gène qui impose le phénotype chez un hétérozygote. S'oppose à récessivité.

Double hélice d'ADN : structure de l'ADN, composée de deux brins en forme d'hélice entrelacés reliés par des liaisons hydrogène par appariement des bases.

Empreinte génétique : aussi appelée « Test ADN », est une technique pour identifier des individus de la même espèce en comparant de courte séquences de leur ADN.

Empreinte parentale : phénomène épigénétique par lequel l'activité d'un gène chez un individu dépend du parent qui l'a transmis. Certains gènes ont donc une empreinte maternelle et d'autres une empreinte paternelle.

Enhancer : stimulateur, défini comme une séquence d'ADN qui, lorsqu'elle est liée par une protéine dite activatrice, facilite l'expression d'un gène en fonction du tissu considéré. Il est parfois très éloigné en amont ou en aval du gène cible.

Épissage : réaction complexe consistant en l'élimination des introns et en l'assemblage des exons constitutifs d'un transcrit primaire d'un gène.

Épissage alternatif : processus par lequel à partir d'un même transcrit primaire des ARN messagers différents sont produits à la suite de variations dans le patron d'épissage du pré-ARNm.

Épistasie : situation dans laquelle l'expression phénotypique différentielle d'un génotype à un locus dépend du génotype à un autre locus.

Exon : toute section non intronique de la séquence codante d'un gène. Lorsqu'ils sont assemblés après l'épissage du transcrit primaire les exons constituent l'ARN messager qui est traduit en protéine.

Épisome : élément génétique qui peut se répliquer de façon indépendante dans le cytoplasme de la bactérie. Il peut également intégrer le chromosome bactérien et se répliquer alors en même temps que lui.

Facteur F : épisome bactérien qui confère à la cellule qui le possède la propriété de transmettre du matériel génétique.

Fréquence allélique : mesure de la présence d'un allèle dans une population par rapport à tous les allèles d'un gène, présents dans la même population.



Fréquence génotypique : proportion d'individus ayant un génotype donné au sein d'une population.

Génome : ensemble du matériel génétique contenu dans le ou les chromosomes d'une cellule.

Génotype : composition allélique spécifique d'une cellule pour un gène, un groupe de gènes voire la totalité de son génome.

Haploïde : cellule ou organisme ne possédant qu'un seul jeu de chromosome.

Fréquence de recombinants : pourcentage ou proportions d'individus ou de cellules recombinants dont le phénotype diffère des phénotypes parentaux de départ.

Génération F1 : première génération de descendants produite par le croisement de deux lignées parentales.

Génération F2 : deuxième génération de descendants issue d'un croisement entre individus d'une même F1 ou par autofécondation chez les végétaux.

Haplotype : classe génétique décrite par une séquence d'ADN (combinaison de SNPs) ou de gènes présents sur un chromosome (combinaison des allèles).

Hérédité : ressemblance des caractères entre parents et descendants.

Héritabilité : proportion de la variance phénotypique dans une population attribuable à la variance génotypique.

Hétérozygote : organisme possédant pour un même gène deux allèles différents.

Histone : protéine basique constituant des complexes octamériques (2 exemplaires de chacune des histones H2A, H2B, H3 et H4) autour desquels l'ADN est enroulé pour former les nucléosomes des chromosomes eucaryotes. L'histone H1 participe à la cohésion de l'ensemble.

Homéoboîte : courte séquence d'ADN d'environ 180 pb, située au sein d'un gène impliqué dans la régulation du développement des animaux, champignons et plantes. Ces gènes sont nommés pour cette raison « gènes à homéoboîtes » et forment la famille des gènes homéotiques.

Homéodomaine : domaine protéique d'environ 60 acides aminés, présent dans la séquence peptidique de l'homéoboîte et capable de se lier à l'ADN.

Homozygote : organisme possédant pour un même gène deux allèles identiques.



Inducteur : molécule ou plus généralement agent de l'environnement qui stimule la transcription de l'ADN à partir d'un promoteur.

Interférence : mesure dans une région chromosomique de l'indépendance des crossing-over les uns des autres.

Intron : région non codante d'un gène eucaryote, située entre deux exons, transcrite en ARN mais ensuite excisée par épissage et absente de l'ARN messager mature traduit en protéine.

IRES : « internal ribosome entry site », séquences internes particulières des ARN messagers eucaryotes et des ARN viraux eucaryotes fonctionnant comme des sites d'entrée des ribosomes à l'instar des sites de liaison RBS situés à l'extrémité 5' de l'ARN messager procaryote.

Isolateur : nom donné à une séquence régulatrice au sein de l'ADN séparant les stimulateurs (enhancers) de certains promoteurs. Lorsqu'il est occupé par une protéine appropriée l'isolateur bloque la communication entre les protéines activatrices liées aux stimulateurs et certains promoteurs.

Jonction de Holliday : structure intermédiaire dans la recombinaison génétique impliquant deux molécules d'ADN double brin, reliées par l'intermédiaire d'un crossing-over réciproque intéressant un brin de chaque molécule.

Kinétochore : complexe de protéines auxquelles s'attache les microtubules constituant les fibres du fuseau mitotique nucléaire jouant un rôle actif dans le mouvement des chromosomes vers les pôles.

Liaison à l'X : mode de transmission des gènes présents dans la région différentielle du chromosome X et absents du chromosome Y.

LINE : appelés longs éléments nucléaires dispersés, ils correspondent à des éléments transposables de classe 1 codant une transcriptase inverse.

Lysogénie : un des deux types de développement s'appliquant à une bactérie infectée par un phage tempéré. Dans ce mode, le génome du phage est réprimé. Il s'intègre dans le génome bactérien et se réplique comme ce dernier. Sous l'effet d'une induction environnementale mettant en cause la survie de la bactérie-hôte, la dérépression du génome viral se produit et le mode de développement lytique s'installe entraînant la production de nouvelles particules virales et leur libération dans l'environnement par la lyse de la bactérie.

Méiose : deux divisions successives qui produisent les gamètes chez les animaux et les spores chez les champignons et végétaux. Les 4 cellules qui en résultent, possèdent la moitié du matériel génétique de la cellule d'origine.



Microsatellite : nom donné à de courtes séquences d'ADN constituées par la répétition en nombre variable selon les individus de di- ou trinucléotides. Exemple : la répétition de nucléotides contenant thymine et guanine, $(TG)_n$.

Microtubule : tubules fins et creux assemblés à partir de sous-unités de la tubuline. Ils forment le plus important des trois systèmes de cablage filamenteux du cytosquelette.

Morphogène : molécule qui intervient directement par son activité et en fonction de sa concentration dans la morphogenèse d'un organisme.

Multiallélisme : traduit un nombre élevé d'allèles différents pour un même gène dans une population donnée.

Mutation ponctuelle : modification héréditaire de la séquence nucléotidique d'un chromosome portant sur une seule base purique (A, G) ou pyrimidique (T, C). Il peut s'agir d'une transversion (Pu ⇔ Py) ou d'une transition (Pu ⇔ Pu ou Py ⇔ Py).

Nucléosome : unité de base de la structure d'un chromosome eucaryote (voir histone).

Octade : asque contenant huit ascospores, produit chez des espèces dont la tétrade subit une division mitotique après la méiose.

Octamère : complexe moléculaire composé de 8 sous-unités ou monomères. Le noyau protéique d'un nucléosome se compose d'un octamère d'histones.

Opérateur : désigne une région de l'ADN situé au début d'un ensemble de gènes adjacents transcrits simultanément (opéron). L'opérateur est le site de liaison d'une protéine dite répresseur de l'opéron.

Organisme transgénique : organisme dont le génome a été modifié par l'introduction d'un ADN étranger porteur de nouveaux gènes.

Origine de réplication : séquence nucléotidique ou site de l'ADN où démarre la réplication.

Paramutation : désigne l'effet d'un des deux allèles présents à une génération et qui affecte de façon héritable l'autre allèle dans les générations ultérieures, même si l'allèle responsable du changement n'est pas lui-même transmis.

Phénotype : forme adoptée par un caractère, liée à un génotype spécifique.

Pili : ou pilus, appendice porté par une bactérie *E. coli* donneuse et servant de pont pour la transmission d'ADN à une bactérie receveuse lors de la conjugaison.



Plasmide : petite molécule d'ADN circulaire extrachromosomique se répliquant de façon autonome. Considéré comme un outil essentiel de l'ingénierie génétique.

Ploïdie : nombre de jeux de chromosomes présents dans une cellule ou un organisme.

Pré-ARNm : molécule d'ARN dite précurseur de l'ARN messager. Désigne le transcrit primaire d'un gène avant les diverses modifications (épissage des introns, éventuel « editing », polyadénylation) conduisant à la formation de l'ARN messager mature, devenu apte pour la traduction en protéine.

Procaryote : organisme constitué d'une seule cellule sans noyau, le matériel génétique étant dans le cytoplasme.

Promoteur : séquence d'ADN en amont de l'extrémité 5' de la séquence d'un gène, constituant le site de liaison des ARN polymérases assurant le démarrage de la transcription.

Prophage : chromosome phagique inséré au sein de l'ADN d'un chromosome bactérien.

Protéine cro : une des principales protéines régulatrices impliquées dans le mode lytique de l'infection d'une bactérie par le phage λ (voir mode lysogène et mode lytique).

Récessivité : ce terme s'emploie pour désigner un allèle, présent chez un hétérozygote, qui ne peut donner un phénotype lorsqu'il est seul ; le phénotype étant imposé par l'allèle dominant.

Recombinaison : terme général désignant tout processus cellulaire qui engendre une nouvelle combinaison génique ou chromosomique non préexistante dans la cellule en question ou dans ses progéniteurs. Désigne aussi le processus similaire qui se produit à la méiose au cours de la formation d'un produit haploïde différent des haploïdes constituant le diploïde méiotique initial.

Recombinaison homologue : recombinaison entre deux molécules d'ADN de séquences similaires, fréquente au cours de la méiose et de la mitose chez les eucaryotes.

Recombinaison interchromosomique : recombinaison résultant d'un assortiment indépendant des chromosomes à la 1re division méiotique.

Recombinaison intrachromosomique : recombinaison résultant d'un crossing-over entre deux loci portés par des chromosomes homologues.

Répresseur : désigne la protéine qui se lie à une séquence régulatrice de l'ADN ou sur l'opérateur d'un gène et qui bloque la transcription.



Répresseur λ : désigne le répresseur spécifique du phage λ qui bloque la transcription des gènes du mode lytique et oriente le processus infectieux du phage λ vers le mode lysogène marqué par l'intégration du génome viral dans le chromosome bactérien sous la forme du prophage.

Shotgun : méthode qui consiste à fragmenter en morceaux de très petite taille l'ADN chromosomique, à cloner les fragments puis à les séquencer directement.

Silencer : extincteur, défini comme une séquence d'ADN qui, lorsqu'elle est liée par une protéine inhibitrice (répresseur), bloque l'expression d'un gène en fonction du tissu considéré. Cette séquence est parfois très éloignée en amont ou en aval du gène cible.

SINE : ou courts éléments nucléaires dispersés, correspondent à des éléments transposables de classe 1 mais ne codent pas une transcriptase inverse, utilise celle des LINE.

Site d'épissage : séquences aux extrémités 5' et 3' d'un intron grâce auxquelles s'effectue l'épissage de l'intron.

Splicéosome : complexe ribonucléoprotéique de grande taille impliqué dans le processus d'épissage (« splicing ») des transcrits primaires d'ARN.

Stimulateur : ou activateur ou « enhancer », séquences d'ADN, sites de liaison de facteurs de transcription agissant positivement sur la transcription des gènes souvent proches de ces séquences.

Suppresseur : mutation qui peut faire disparaître l'effet d'une mutation sur un autre gène et restituer le phénotype sauvage.

Télomérase : enzyme qui utilise comme matrice une petite séquence d'ARN pour catalyser l'addition d'unités nucléotidiques répétées aux extrémités des chromosomes linéaires eucaryotes afin de compenser leur raccourcissement consécutif à la réplication.

Télomère : séquence particulière située à l'extrémité d'un chromosome linéaire eucaryote.

Topoisomérase : enzyme qui introduit des surenroulements positifs ou négatifs dans un duplex circulaire fermé. Elle clive les polynucléotides des duplex surenroulés afin de relâcher les contraintes puis reconstitue la liaison phosphodiester.

Traduction : processus au cours duquel l'information génétique contenue dans une molécule d'ARN messager détermine la séquence des acides aminés lors de la synthèse protéique.



Transcription : processus enzymatique au cours duquel l'information génétique contenue dans une molécule d'ADN simple brin détermine par complémentarité des bases la séquence en nucléotides d'une chaîne d'ARN.

Transcrit primaire : ARN initial résultant du processus de la transcription (voir pré-ARNm).

Transduction : transfert de gène d'une bactérie à une autre par l'intermédiaire d'un phage.

Transduction généralisée : capacité de certains phages à transduire n'importe quel gène dans le chromosome bactérien.

Transduction spécialisé : capacité pour un phage spécifique de transduire que des régions particulières d'un chromosome bactérien.

Transformation : modification dirigée d'un génome par introduction dans la cellule-hôte d'un ADN exogène ce qui donne à la cellule un nouveau phénotype.

Transition : désigne les mutations ponctuelles substituant entre elles les bases puriques ou les bases pyrimidiques (voir mutation ponctuelle).

Transposition : ensemble des mécanismes moléculaires assurant la mobilité d'un gène ou d'un ensemble de gènes d'un site du génome à un autre.

Transversion : désigne les mutations ponctuelles substituant les bases puriques et pyrimidiques entre elles (voir mutation ponctuelle)

Zygote : cellule formée par la fusion d'un gamète mâle (spermatozoïde) et d'un gamète femelle (ovule). Le diploïde formé se divisera pour créer un organisme diploïde différencié.


sba-medecine
La communauté médicale des étudiants en médecine



Index

Pour se repérer plus facilement, les occurrences principales des mots de l'index sont en gras dans le cours.

A

B

C


sba-medecine
La communauté médicale des étudiants en médecine

D

E

F

G

H

I



S

T

U

V

X